

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-265808

(43)Date of publication of application : 28.09.2001

(51)Int.Cl.

G06F 17/30

(21)Application number : 2000-081037 (71)Applicant : SKYSOFT INC

(22)Date of filing : 22.03.2000 (72)Inventor : HIROOKA YASUO
TERANO TAKAO
OTSUKA YUKICHI

(54) SYSTEM AND METHOD FOR INFORMATION RETRIEVAL

(57)Abstract:

PROBLEM TO BE SOLVED: To search for and recommend a book which is both useful and unexpected to a customer.

SOLUTION: Key word vectors of respective books that the purchaser bought in the past are put together by categories to generate a customer profile. According to a book database, the key word vectors are generated for each book. Then the similarity between the key word vectors of the customer profile and the key word vectors of unpurchased books is computed by combinations of books in the same category and between different categories, and the a book can be retrieved from the point of view of usefulness and unexpectedness. Thus, an unpurchased book with high similarity can be recommended.

LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision
of rejection]

[Kind of final disposal of application other
than the examiner's decision of rejection
or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]
[Number of appeal against examiner's
decision of rejection]
[Date of requesting appeal against
examiner's decision of rejection]
[Date of extinction of right]

CLAIMS

[Claim(s)]

[Claim 1] A classification means to be the information retrieval system with which the contents which a user desires out of the information space containing much contents are searched, and to classify each contents on said information space into the contents category of a predetermined number, A feature-vector generation means to generate the feature vector of each contents on said information space, A user profile generation means to generate the user profile containing the feature vector of the contents in which the user did past access, Information retrieval system characterized by providing a vector comparison means to compare the feature vector of a user profile and contents for every contents category, and a contents specification means to specify the contents which a user desires based on the result of said comparison.

[Claim 2] It is the information retrieval system according to claim 1 characterized by generating a keyword vector with the coordinate value which said contents are the contents containing text documents, such as books, and hung the weighting factor at the frequency of occurrence of each keyword in the inside of contents on N dimension keyword space where said feature-vector generation means assigned the axis of coordinates to each of the keyword of N individual.

[Claim 3] Said user profile generation means is information retrieval system according to claim 1 characterized by generating the user profile which a user compounds the feature vector of each contents which carried out past access for every contents category, and becomes from the feature vector for every contents category.

[Claim 4] It is the information retrieval system according to claim 1 which said user profile generation means generates the user profile which consists of a feature vector for every contents category, and is characterized by said vector comparison means comparing the feature vector of a user profile with the feature vector of contents within the same category.

[Claim 5] It is the information retrieval system according to claim 1 which said user profile generation means generates the user profile which consists of a feature vector for every contents category, and is characterized by said vector comparison means

comparing the feature vector of a user profile with the feature vector of contents between different categories.

[Claim 6] Said contents specification means is information retrieval system according to claim 1 characterized by specifying as contents which a user expects the high contents of a user profile and similarity.

[Claim 7] Said contents specification means is information retrieval system according to claim 1 characterized by a user specifying among the high contents of a user profile and similarity as contents which a user expects a non-accessed thing.

[Claim 8] The feature-vector generation step which is the information retrieval approach of searching the contents which a user desires out of the information space containing much contents, and generates the feature vector of each contents on said information space, The classification step which classifies each contents on said information space into the contents category of a predetermined number, The user profile generation step which generates the user profile containing the feature vector of the contents in which the user did past access, The information retrieval approach characterized by providing the vector comparison step which compares the feature vector of a user profile and contents for every contents category, and the contents specification step which specifies the contents which a user desires based on the result of said comparison.

[Claim 9] said contents be the information retrieval approaches according to claim 8 which be the contents containing text documents , such as books , and be characterize by generate a keyword vector with the coordinate value which hung the weighting factor at the frequency of occurrence of each keyword in the inside of contents on N dimension keyword space which assigned the axis of coordinates to each of the keyword of N individual at said feature vector generation step .

[Claim 10] The information retrieval approach according to claim 8 characterized by generating the user profile which a user compounds the feature vector of each contents which carried out past access for every contents category, and becomes from the feature vector for every contents category at said user profile generation step.

[Claim 11] The information retrieval approach according to claim 8 which generates the user profile which consists of a feature vector for every contents category at said user profile generation step, and is characterized by comparing the feature vector of a user profile with the feature vector of contents within the same category at said vector comparison step.

[Claim 12] It is the information retrieval approach according to claim 8 which generates the user profile which consists of a feature vector for every contents category at said user profile generation step, and is characterized by said vector comparison step comparing the feature vector of a user profile with the feature vector of contents between different categories.

[Claim 13] The information retrieval approach according to claim 8 characterized by

specifying as contents which a user expects the high contents of a user profile and similarity at said contents specification step.

[Claim 14] The information retrieval approach according to claim 8 characterized by a user specifying among the high contents of a user profile and similarity at said contents specification step as contents which a user expects a non-accessed thing.

[Claim 15] It is the software storage which stores the computer software for making the information retrieval processing which searches the contents which a user desires out of the information space containing much contents perform on computer system in a computer-readable format. The feature-vector generation step to which said computer software generates the feature vector of each contents on said information space, The classification step which classifies each contents on said information space into the contents category of a predetermined number, The user profile generation step which generates the user profile containing the feature vector of the contents in which the user did past access, The software storage characterized by providing the vector comparison step which compares the feature vector of a user profile and contents for every contents category, and the contents specification step which specifies the contents which a user desires based on the result of said comparison.

[Claim 16] Said contents be software storages according to claim 15 which be the contents containing text documents , such as books , and be characterize by generate a keyword vector with the coordinate value which hung the weighting factor at the frequency of occurrence of each keyword in the inside of contents on N dimension keyword space which assigned the axis of coordinates to each of the keyword of N individual at said feature vector generation step .

[Claim 17] The software storage according to claim 15 characterized by generating the user profile which a user compounds the feature vector of each contents which carried out past access for every contents category, and becomes from the feature vector for every contents category at said user profile generation step.

[Claim 18] The software storage according to claim 15 which generates the user profile which consists of a feature vector for every contents category at said user profile generation step, and is characterized by comparing the feature vector of a user profile with the feature vector of contents within the same category at said vector comparison step.

[Claim 19] It is the software storage according to claim 15 which generates the user profile which consists of a feature vector for every contents category at said user profile generation step, and is characterized by said vector comparison step comparing the feature vector of a user profile with the feature vector of contents between different categories.

[Claim 20] The software storage according to claim 15 characterized by specifying as contents which a user expects the high contents of a user profile and similarity at said contents specification step.

[Claim 21] The software storage according to claim 15 characterized by a user specifying among the high contents of a user profile and similarity at said contents specification step as contents which a user expects a non-accessed thing.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Field of the Invention] This invention relates to the information retrieval system and the approach for discovering the contents which agreed in the demand of a user out of huge information contents, and relates to the information retrieval system and the approach of searching for desired contents especially based on user profile information, such as a hobby and taste of a user, and the past retrieval hysteresis.

[0002] Furthermore, in order RIKOMENDO [the goods which a customer desires when this invention performs business, such as a goods sale, on the Internet in detail / discover and], available information retrieval system and an available approach are started, and it is especially related with the information retrieval system and the approach which are applied in goods sale business which asks for the goods with which a customer has benefit and unexpected nature like a books sale.

[0003]

[Description of the Prior Art] It has become possible with improvement in the technology of the semi-conductor manufacture and others in these days to offer comparatively cheap computing systems including a personal computer (PC) in a commercial scene. A computing system is spreading extensively until it reaches the office and the ordinary homes of an enterprise or research facilities. Furthermore, the want to the network computing technique which interconnects in computer systems has also been mounting. On a network, computer resources, such as a file and a printer, can be shared between each users, or the collaboration-activity of informational share, negotiation, distribution, exchange, etc. can be done smoothly.

[0004] The gestalt of the network which interconnects in computers is various. For example, LAN (Local Area Network) laid by part like Ethernet (trademark) or a token ring, WAN (Wide Area Network) which consists of dedicated lines etc. by interconnecting in LANs, the "Internet" (The Internet) which accomplished growth to the network of worldwide magnitude literally as a result of having repeated interconnect of networks and having performed it further, etc. are various.

[0005] The Internet came to grow up to be a huge network, as a result of the server installed in each university, research facilities, etc. repeating interconnect independently. The servers on the Internet usually interconnect with the TCP/IP

(Transmission Control Protocol/Internet Protocol) base.

[0006] On the Internet, service of a large number which operate on TCP/IP networks, such as WWW (World Wide Web), News, and TELNET (TELEtypewriter NETwork), FTP (File Transfer Protocol), Gopher, is exhibited. Especially WWW is broader-based information retrieval system which offers the information space of hyperlink structure, and has also become the greatest factor which accomplishes the explosive growth and the rapid spread of the Internet.

[0007] WWW is built with a client-server mold. A WWW server is computer system which works the server software called a HTTP demon, and stores the information resource described in the HTML format, and/or offers this resource. A WWW server exhibits information resources including a HTML document by onerous or onerous on the Internet. If TCP/IP is followed, these information resources are specified by the identifier of the format of URL (Uniform Resource Locator). Although a HTML document can include link relation with other information resources, a link information is also specified in a URL format.

[0008] The WWW client of another side is a computer which operates using the user agent called a "WWW browser" or a "HTML browser." "Netscape Navigator" of U.S. Netscape Communications and "Internet Explorer" of U.S. Microsoft are the examples of representation of a WWW browser.

[0009] A WWW client user can specify a desired information resource by entering the URL character string as a resource identifier in the location box on a browser screen. A WWW browser usually performs access to the information resource on a WWW server, i.e., the ejection of an information resource, according to a HTTP (Hyper Text Transfer Protocol) protocol. Furthermore, a WWW browser analyzes the acquired HTML document and displays the content on a browser window screen (that is, a homepage is assembled on a browser screen). On the WWW page displayed on the browser screen, a link information is displayed as "support." By clicking support, a user can do sequential assignment of the information resource which is the following reference place. That is, a user can search the WWW information-resource space around which the hyperlink was spread on the Internet through a WWW browser screen, and can access various information resources without a joint one after another (common knowledge). WWW is equipped with the advanced user interface which can perform information retrieval by the easy actuation by the dialogue input.

[0010] The diffusion rate of the Internet is high and has already permeated deeply various research facilities, such as a university, the office of an enterprise, ordinary homes, etc. Current and the Internet are used as a means of not only mere information distribution but various business. They are the commercial distribution and the sale system called the so-called "internet business."

[0011] For example, digital information, such as computer software, and music, image contents, can be distributed via the Internet. In such a case, most of the distribution channel and sale way of the goods needed in case it sells by archive media, such as

CD and DVD, is omissible.

[0012] Moreover, when internet business is applied to the conventional mail order, goods information can be momentarily distributed all over the world only by uploading the data contents equivalent to a catalog on one website. Consequently, the time and effort and cost which distribute or mail print media, such as a throwaway and a catalog, can be saved. There is instancy nature of goods information and commitment of goods can be concluded smoothly and promptly.

[0013] By distributing goods information on the Internet, the consumer or commodity demand buried in the remote place can be unearthed comparatively easily. For a goods vendor side, even if it is little goods with which the shipment high to the extent that it arranges to a shelf space is not expected, it can sell at pinpoint to the consumer who are scattered in various places. Moreover, if it carries out from a consuming public side, even if it is goods which are not dealt with at a nearby firm but have feeling of a comparatively high price, such as an importation and a foreign book, a supplier company can be discovered easily and it can purchase by low cost. In other words, the threshold of a store becomes markedly low.

[0014] Internet business is excellent in informational instancy nature. Moreover, a physical means like the showcase for a goods sale as other advantages is hardly needed, but it is mentioned that the initial cost which a goods sale and sales promotion take is low. By development of internet business, a consumer's position becomes strong and the width of face of goods selection is said to have been extended greatly. furthermore, the thing of a tariff which it pays and is alike and is adopted for secure electronic banking -- a goods sale -- all processes can almost be mechanized and automated by remote operation via a network.

[0015] Fundamentally in internet business, it can trade through being interactive and an advanced user interface like WWW. Moreover, in the enterprise which develops internet business, it aims at PASONA rise-ization which customizes contents and service for every customer taking advantage of the point of being interactively connected to the customer.

[0016] However, while it is realizable with the existing technique in WWW or an electronic mail, it is not easy to perform the response according to individual for every customer to prepare suitable contents for every customer, or to recommend. To such a technical problem, the "RIKOMENDA system" which generates recommendation information automatically is expected.

[0017] For example, while being able to offer the new title according to liking to each reader and expanding customer service by applying a RIKOMENDA system to business which performs a books sale of a foreign book etc. on the Internet, it leads also to buildup of sales.

[0018] Two methods called a "contents analysis method" and a "coordination filtering method" exist in the RIKOMENDA system conventionally.

[0019] A contents analysis method is the approach, RIKOMENDO [choose near

contents by a user's interest and], by comparing the user profile which is the description expression of contents and user interest expression used as a RIKOMENDO object. When applying to a foreign-book sale, the description expression of contents is accumulated as a books database which consists of the title of each books, an author name, a genre, a keyword, etc. Moreover, a user profile consists of the books purchasing hysteresis and retrieval hysteresis of the past besides profiles, such as a user's hobby and special ability.

[0020] For contents analysis, the common knowledge technique of the information retrieval field is applicable. For example, vector-space model by the term with weight

[1] Tools of analysis which compare the feature vectors about contents like retrieval of a mold can be used. That is, it searches for, the vector, i.e., a "keyword vector", which has the coordinate value which hung the weighting factor on the count of an appearance of each keyword in the inside of contents on N dimension space which assigned the axis of coordinates to each of the keyword of N individual, and desired contents are searched based on the similarity of keyword vectors. It can ask for similarity based on the inner product and distance of a vector.

[0021] However, two next points are pointed out as a fault of a contents analysis method [2].

(1) The contents made applicable to retrieval are dependent on contents analytical skill. In the actual condition, it is restricted to the text document which can apply an information retrieval technique.

[0022] (2) There is an inclination only for information similar to the information which the user esteemed to be extracted.

[0023] When using information retrieval for RIKOMENDESHON of books, such as a foreign book, especially, only the things of the books purchased before and a similar content are not always liked. First of all, since reading is an act based on intellectual desire, it is desirable there is not only benefit, such as to agree to hobby and taste of a user (reader), but its RIKOMENDO [books with unexpected nature or freshness]. If RIKOMENDESHON is repeated only in about the books of a similar content, a customer will get bored soon. Therefore, the fault of the above (2) which a contents analysis method has may become fatal when applying to books sale business.

[0024] Moreover, a coordination filtering method is a method which generates RIKOMENDO information by analyzing the contents assessment information between two or more users, without performing contents analysis. First, grouping of the associate with a similar interest is carried out by comparing the contents assessment information between users. And RIKOMENDO [the thing whose oneself is not evaluated] among the contents which his associate esteemed. The coordination filtering method is already adopted in some commercial solutions represented by NetPerceptions [3] and [4].

[0025] In order not to perform contents analysis, a coordination filtering method has the advantage that there is no limit in the target contents, and can be applied also to

almost all media other than a text document. On the other hand, four next points are pointed out as a fault of a coordination filtering method [2].

[0026] (1) The sufficient number of users and an assessment information number are needed to the amount of object contents. When the numbers of users run short, coverage becomes narrow and it becomes impossible to make a RIKOMENDO object only the part restricted among all contents. Moreover, a user with a unique taste is anxious about RIKOMENDO itself becoming impossible etc. as a result in which an associate is not found.

[0027] (2) Even if new contents are registered, it does not become a RIKOMENDO object until someone evaluates.

[0028] (3) Since contents analysis is not performed, although the content is similar, assessment of the contents from which ID differs cannot make high users an associate.

[0029] (4) It is difficult for the incorrect input of contents assessment to eliminate the effect which it has on RIKOMENDO to other users.

[0030] When RIKOMENDO service is considered for books, fundamentally, both a contents method and a coordination filtering method can be applied. However, since an all directions type has a fault, respectively as mentioned above, we are anxious about the following point.

[0031] (1) Possibility that the books [RIKOMENDO / books] will become only things similar to the books which the customer purchased in the past by the contents analysis method is high. If RIKOMENDESHON is repeated only in about the books of a similar content, a customer will get bored soon.

[0032] (2) By books sale, since the target number of contents is huge, if the number of customers set as the object of RIKOMENDO service is not considerable magnitude, either, high quality service is not expectable with a coordination filtering method.

[0033]

[Problem(s) to be Solved by the Invention] The object of this invention is to offer the outstanding information retrieval system and the outstanding approach of RIKOMENDO [the contents corresponding to a demand of a user / discover and] out of huge information contents.

[0034] The further object of this invention is to offer the outstanding information retrieval system and the outstanding approach of desired contents that it can search and RIKOMENDO can be performed, based on user profile information, such as a hobby and taste of a user, and past retrieval hysteresis, purchasing hysteresis.

[0035] The further object of this invention is to offer the outstanding information retrieval system and the outstanding approach of RIKOMENDO [the goods which a customer expects / discover and], when performing business, such as a goods sale, on the Internet.

[0036] The further object of this invention is to offer the outstanding information retrieval system and the outstanding approach of RIKOMENDO [goods with which a

customer asks for the unexpected nature other than benefit / discover and] like a books sale.

[0037]

[Means for Solving the Problem] This invention is made in consideration of the above-mentioned technical problem. The 1st side face The classification means or step which is the information retrieval system or the approach of searching the contents which a user desires out of the information space containing much contents, and classifies each contents on said information space into the contents category of a predetermined number, The feature-vector generation means or step which generates the feature vector of each contents on said information space, The classification means or step which classifies each contents on said information space into the contents category of a predetermined number, The user profile generation means or step which generates the user profile containing the feature vector of the contents in which the user did past access, The vector comparison means or step which compares the feature vector of a user profile and contents for every contents category, It is the information retrieval system or the approach characterized by providing the contents specification means or step which specifies the contents which a user desires based on the result of said comparison.

[0038] However, in the information retrieval system or the approach concerning this invention, when dealing with the goods with which the contents category is beforehand specified like books as contents for retrieval, an above-mentioned classification means or an above-mentioned step may be skipped, and the existing contents category may be used as it is. An example of a books category is "BISAC Subject Category" which BISAC (BookIndustry Systems Advisory Committee) offers.

[0039] In the information retrieval system or the approach concerning the 1st side face of this invention, said contents may be the contents containing text documents, such as books. In such a case, you may make it said feature-vector generation means or step generate the keyword vector which has the coordinate value which hung the weighting factor on the frequency of occurrence of each keyword in the inside of contents on N dimension keyword space which assigned the axis of coordinates to each of the keyword of N individual.

[0040] Moreover, said user profile generation means or step compounds the feature vector of each contents in which the user did past access for every contents category, and you may make it generate the user profile which consists of a feature vector for every contents category.

[0041] Moreover, in said user profile generation means or step, when generating the user profile which consists of a feature vector for every contents category, even if like [said vector comparison means or step / compare the feature vector of a user profile with the feature vector of contents and] (order matching) within the same category, it is good. Or even if like [said vector comparison means or step / compare the feature vector of a user profile with the feature vector of contents and]

(crossover matching) between different categories, it is good.

[0042] Moreover, said contents specification means or step is specified as contents which a user expects the high contents of a user profile and similarity, and it may be made RIKOMENDO [it / a user].

[0043] Moreover, among the high contents of a user profile and similarity, a user specifies said contents specification means or step as contents which a user expects a non-accessed thing, and may be made RIKOMENDO [it / a user].

[0044] Moreover, the 2nd side face of this invention It is the software storage which stores the computer software for making the information retrieval processing which searches the contents which a user desires out of the information space containing much contents perform on computer system in a computer-readable format. The feature-vector generation step to which said computer software generates the feature vector of each contents on said information space, The classification step which classifies each contents on said information space into the contents category of a predetermined number, The user profile generation step which generates the user profile containing the feature vector of the contents in which the user did past access, The vector comparison step which compares the feature vector of a user profile and contents for every contents category, It is the software storage characterized by providing the contents specification step which specifies the contents which a user desires based on the result of said comparison.

[0045]

[Function] The information retrieval system and the approach concerning this invention are applicable suitable for retrieval, RIKOMENDESHON, etc. of contents which consist of text documents, such as books.

[0046] By the information retrieval system and the approach concerning this invention, a customer compounds the feature vector of each books purchased in the past according to a category, and generates the customer profile which consists of a feature vector for every category. Moreover, a feature vector is generated for every books by performing keyword vector-space configuration processing to each books based on a books database.

[0047] And RIKOMENDO [the similarity of the feature vector of a customer profile and the feature vector which each non-purchased books have is calculated, and / the high non-purchased books of similarity]. Books can be searched with performing similarity count of a feature vector in each combination between categories which are different in a list (crossover matching) from a viewpoint of the both sides of benefit and unexpected nature between the same categories (order matching).

[0048] The feature vector dealt with in this invention is good by the "keyword vector" expressing the frequency of occurrence of the keyword which are scattered in contents, such as for example, a text document. A keyword vector is a vector which has the coordinate value which hung the weighting factor on the count of an appearance of each keyword in the inside of contents on N dimension keyword space

which assigned the axis of coordinates to each of the keyword of N individual.

[0049] Although the information retrieval system and the approach concerning this invention make a contents analysis method the keynote, they can cancel the fault of a contents analysis method by carrying out the device which expands the width of face of RIKOMENDO information. Moreover, also when customer magnitude is small, offer of high quality RIKOMENDO service can be enabled because the number of object contents applies this invention in a very big type of industry like a books sale.

[0050] The software storage concerning the 2nd side face of this invention is a medium which offers computer software in a computer-readable format to the general purpose computer system which can perform various program codes, for example. Attachment and detachment of CD (Compact Disc), FD (Floppy Disc), MO (Magneto-Optical disc), etc., etc. are free for such a medium, and it is a storage of portability.

[0051] Such a software storage defines the collaboration-relation on the structure of the computer software and the storage for realizing the function of computer software predetermined in a computer system top, or a function. If it puts in another way, by installing predetermined computer software in computer system through the software storage concerning the 2nd side face of this invention, on computer system, a collaboration-operation is demonstrated and the same operation effectiveness as the information retrieval system or the approach concerning the 1st side face of this invention can be acquired.

[0052] The object, the description, and advantage of further others of this invention will become [rather than] clear by detailed explanation based on the example and the drawing to attach of this invention mentioned later.

[0053]

[Embodiment of the Invention] The information retrieval system concerning this invention can be applied to for example, a books sale, can discover the books which each customer desires from the crest of the books of a huge number, and can recommend RIKOMENDO, i.e., purchasing.

[0054] In drawing 1 , the procedure in which the information retrieval system concerning this invention generates RIKOMENDO information is illustrated.

[0055] Although it is common in contents analysis type RIKOMENDESHON for a text document to use a keyword vector as a description expression of the customer profile which is the expression of customer interest, or contents, the same is said of the case of the information retrieval system concerning this invention. A keyword vector is N dimension vector which has the coordinate value which hung the weighting factor on the count of an appearance of each keyword in the inside of contents on N dimension space which assigned the axis of coordinates to each of the keyword of N individual (above-mentioned).

[0056] In this information retrieval system, the keyword vector of a customer profile is generated using books information (books database) and a customer's purchasing hysteresis information (purchasing hysteresis database). Here, the keyword vector of

a customer profile consists of vectors by which a customer compounds the keyword vector of each books purchased in the past according to a category, and is generated. Purchasing hysteresis information consists of a customer's member identifier (ID), an identifier (ISBN:International Standard BookNumber) of the books purchased in the past, etc.

[0057] On the other hand, a keyword vector is generable for every books by performing keyword vector-space configuration processing to each books based on a books database. A books database consists of the title of each books, content explanation (an epitome/abstract), an author name, a category, etc.

[0058] And the similarity of the keyword vector of a customer profile and the keyword vector of each non-purchased books is calculated. RIKOMENDO [the high non-purchased books of similarity] to a customer.

[0059] Subsequently, it explains in detail about creation processing of a customer profile.

[0060] In this invention, the action which discovers the books whose interest the customer suited at the actual bookstore assumes following a books purchasing behavioral model as shown in drawing 2 . Action until it results in the decision of purchase books is divided roughly into two of the actions which choose the action which chooses the bookshelf which suited the interest field, and the books which met interest using the corresponding information on the back cover of the books of a bookshelf, or a table of contents as a graphic display. From the description of each action, "classification mold selection action", a call, and the latter can be called "retrieval mold selection action" for the former.

[0061] In the information retrieval system concerning this invention, two kinds of customer interest expression systems called the keyword vector used as the books category equivalent to classification mold selection action and the criteria of retrieval selection action are adopted based on the books purchasing behavioral model shown in drawing 2 .

[0062] You may build from zero for the information retrieval system concerning this invention, and a books category can also apply the existing books category. As an existing books category, "BISAC Subject Category" which BISAC (Book Industry Systems Advisory Committee) offers can be mentioned, for example. BISAC SubjectCategory It has classified into 49 kinds of categories, such as Art, Business & Economics, Cooking, and Computers.

[0063] In drawing 3 , it is BISAC. The graph which investigated a customer's number of books purchasing categories is shown using Subject Category. Ranging over two or more categories, he has purchased books in many cases like the person who purchased two or more books, and could understand that there is an inclination for the width of face of the interest over common books to be very wide.

[0064] When performing books sale business on the Internet in which various kinds of information retrieval system including WWW (World Wide Web) was substantial, it is

common that a user, i.e., a customer, accesses a sale site himself, and he searches the target books. A books title, an author name, a publishing company, a keyword, etc. are used for retrieval. In the case of the customer who is interested in two or more categories, what a completely different keyword for every books category to search will be properly used for can be imagined easily.

[0065] Therefore, when creating the customer profile expressing customer interest, and the width of face of a customer's interest is wide, it is considered that it is necessary to create two or more keyword vectors.

[0066] Two or more keyword vectors according to purchasing books category constituted the customer profile from the information retrieval system concerning this invention, assuming a books category to be the unit of customer interest.

[0067] A keyword vector is used as a concrete customer profile. A keyword vector is a vector which has the coordinate value which hung the weighting factor on the count of an appearance of each keyword in the inside of contents on N dimension keyword space which assigned the axis of coordinates to each of the keyword of N individual (above-mentioned).

[0068] A keyword vector consists of items on a books database, using a title, content explanation, an author, and a subcategory (what classified 49 classifications into the detail further about 3,700) as an item representing the description of books. Although the phrase in the data of a title or the text format of content explanation is used as a keyword, you may make it also include the information on an author or a subcategory into a vector as one keyword generally.

[0069] At this example, a customer profile consists of keyword vectors for every books category. In drawing 4 , the keyword vector creation procedure of a customer profile is illustrated. First, a keyword vector is given for every books. Subsequently, when this customer adds together the keyword vector of each books purchased in the past for every books category, the keyword vector for every purchasing books category of a customer is created.

[0070] Moreover, to drawing 5 , the creation procedure of the keyword vector for every books is illustrated. However, it is necessary to perform a text manipulation about the title of books, and content explanation. The content of the text manipulation is the following procedure.

[0071] (1) Give a part of speech to each word contained in a text. As a tool for giving a part of speech, it is Brill's English tagger. [5] is used.

[0072] (2) Deletion of an unnecessary word deletes parts of speech a noun, an adjective, and other than a foreign word using part-of-speech information. Then, the unnecessary word list created in advance is applied.

[0073] (3) Use the word which remained without being deleted as a word keyword.

[0074] (4) A compound keyword uses the combination as a compound keyword, when a noun continued and appears in a text.

[0075] About an author, a books category, and a subcategory, weighting is performed

like an above-mentioned word keyword and an above-mentioned compound keyword, and it adds to a keyword vector. In this example, the weighting technique which used the frequency of occurrence TF of a phrase (Term Frequency) and the reverse reference frequency IDF (Inverse Document Frequency) and which is called TFIDF [1] is used for the approach of weighting.

[0076] Let the number of books in which the frequency of occurrence of the j-th keyword t_j of the word, the compound, author, and subcategory in the i-th books d_i is set to f_{ij} , and Keyword t_j appears in the number N of complete-works family registers be what is set with n_j in the following explanation. Moreover, weight set up for every classification of a keyword is set to k. In this description, experientially, weight for every keyword classification was set to $k=2$ about the subcategory, and, other than this, was boiled, therefore was set to $k=1$.

[0077] Moreover, in order to ease the effect of [when the same keyword appears superfluously], we decided to evaluate the frequency of occurrence of each keyword using TF_{ij} defined by the formula of the following which took in pair number representation, not using the frequency of occurrence f_{ij} of the j-th keyword t_j in the i-th books d_i as it is.

[0078]

[Equation 1]

$$TF_{ij} = \frac{f_{ij}}{\sum_{j=1}^n f_{ij}}$$

[0079] Moreover, we decided to take in pair number representation and to use IDF_j defined by the following formulas about reverse reference frequency.

[0080]

[Equation 2]

$$IDF_j = \frac{1}{\sum_{i=1}^n \frac{1}{f_{ij}}}$$

[0081] Moreover, the weight w_{ij} given to the keyword t_j in Books d_i serves as a degree type using each above-mentioned type [a-one number] and above-mentioned [-two number].

[0082]

[Equation 3]

$$w_{ij} = \frac{TF_{ij} \cdot IDF_j}{\sum_{j=1}^n TF_{ij} \cdot IDF_j}$$


[0083] The final keyword vector for every books gives the keyword of 12 high orders to the descending order of the weight w_j of Keyword t_j at each books. However, when realizing this invention, there is not necessarily no need of setting up the upper limit of a keyword, and a upper limit is not limited to 12 pieces, either.

[0084] As shown in drawing 4 , a customer profile consists of the sums for every category of the keyword vector of the purchasing books for every books category, i.e., a synthetic vector. Weight w'_{jkl} of the keyword t_j about Customer's U_k category Cl is

defined like a bottom type.

[0085]

[Equation 4]



[0086] However, in a top type [a-four number], bik sets to 1, when Customer Uk has purchased Books di, and it is set to 0 except it. Moreover, sil sets to 1, when Books di belong to Category Ci, and it is set to 0 except it.

[0087] Selection of the books [RIKOMENDO / books / a customer] is performed by calculating the similarity showing the keyword vector of a customer profile, and the description of books of a keyword vector. The cosine of the angle between keyword vectors to make is used for the count approach of similarity.

[0088] If the weight of the keyword tj in Customer's Uk category Ci is set with w'jkl and it sets with the weight wij of the keyword tj in Books di, rikl will become as a degree type whenever [recommendation / of the books di to Customer's Uk category Ci].

[0089]

[Equation 5]



[0090] Subsequently, RIKOMENDESHON processing of books is explained.

[0091] Generally in the contents tools of analysis using a feature vector like a keyword vector, the description comparison is performed based on the similarity of vectors.

[0092] In this example, the description comparison is performed for every category. Two kinds, "order matching type RIKOMENDESHON" which more specifically compares the keyword vectors contained in the same books category, and "crossover matching type RIKOMENDESHON" which compares the keyword vectors contained in a different books category, are performed.

[0093] Order matching type RIKOMENDESHON procedure is illustrated in drawing 6 .

[0094] It consists of synthetic vectors of the keyword vector in each books category with the past purchasing hysteresis, a customer's interest expression, i.e., customer profile. In the example shown in drawing 6 , the keyword vector expressing a customer profile is generated in each of the books categories A and B which the customer had purchased in the past.

[0095] On the other hand, the keyword vector of the books contained in each of the books categories A and B to which the keyword vector for every books is generated (refer to drawing 4), among these a customer's purchasing hysteresis was left behind is extracted based on a books database.

[0096] Furthermore, it calculates within the same books categories A and B, a comparison, i.e., the similarity, of the keyword vector of a customer profile, and the keyword vector of books. What is necessary is just to use the above-mentioned [-five number] for count of similarity. And RIKOMENDO [a customer / a customer] by making a non-purchased thing into a retrieval result among the high books of similarity.

[0097] In order matching type RIKOMENDESHON, only by saying [that some keywords were in agreement], will be RIKOMENDO the books of the category which is completely uninterested, namely, they can prevent unnecessary or useless RIKOMENDESHON.

[0098] Moreover, crossover matching type RIKOMENDESHON procedure is illustrated to drawing 7 .

[0099] It consists of synthetic vectors of the keyword vector in each books category with the past purchasing hysteresis, a customer's interest expression, i.e., customer profile. In the example shown in drawing 7 , the keyword vector expressing a customer profile is generated in each of the books categories A and B which the customer had purchased in the past. Moreover, the keyword vector of the books contained in each of the books categories A and B to which a customer's purchasing hysteresis was left behind is extracted based on a books database (same as the above).

[0100] Furthermore, the keyword vector of a customer profile is compared with the keyword vector of books between different books categories. In the example shown in drawing 7 , while comparing the customer profile of the books category A with the books keyword vector of the books category B, the customer profile of the books category B is compared with the books keyword vector of the books category A. What is necessary is just to use the above-mentioned [-five number] for the comparison of a vector, i.e., count of similarity. And RIKOMENDO [a customer / a customer] by making a non-purchased thing into a retrieval result among the high books of similarity.

[0101] According to crossover matching type RIKOMENDESHON, the high RIKOMENDO information on unexpected nature is generable by making the relation of the books category of a keyword vector and the books description cross.

[0102] For example, when performing action which looks for the books which have a required keyword from the target bookshelf like the books retrieval phase of a books purchasing behavioral model shown in drawing 2 , although it is interested from usually, it is the case where the books which completely contain the keyword of another field are found by chance. Crossover matching type RIKOMENDESHON is the technique of aiming at such effectiveness intentionally.

[0103] As an intelligible example, RIKOMENDESHON to the customer who has purchased some books is considered from the books and the Cooking category about the category of Science Fiction to Star Trek. According to crossover matching type RIKOMENDESHON, the books of a Cooking category can be searched using keyword

"Star Trek", and books called "Star Trek Cookbook" can be introduced.

[0104] However, in crossover matching, it is considered when limiting to the books category in which had a purchase track record in the past in the books category set as the comparative object, and the customer is interested improves RIKOMENDESHON precision, are desirable.

[0105]

[Example] Hereafter, the example of this invention is explained in detail, referring to a drawing.

[0106] The hardware configuration of the information retrieval system 10 with which operation of this invention is presented is typically shown in drawing 8. Information retrieval system 10 is realized also by performing application (it considering as "RIKOMENDO application" temporarily in this description) predetermined in a common computing system top, although you may design as hardware of dedication. An example of a computing system is the PC / AT compatible machine or its succeeding machine of U.S. IBM. Hereafter, each part of a system 10 is explained.

[0107] CPU (Central Processing Unit)11 which is the Main controller of a system 10 performs various kinds of applications under control of an operating system (OS). OS is good at Windows 98 of UNIX (trademark) or U.S. Microsoft Corp./NT.

[0108] CPU11 interconnects with other equipments by bus 19 as the graphic display. The memory address or I/O Address of a proper is given to each device on a bus 19, respectively, and device access is possible for CPU11 by these addresses. An example of a bus 19 is a PCI (Peripheral Component Interconnect) bus.

[0109] RAM (Random Access Memory)12 is an volatile store used since the program code performed in CPU11 is loaded or the activity data under activation are stored temporarily, and usually consists of two or more DRAM (dynamic RAM) chips. Moreover, ROM (Read Only Memory)13 is the storage of a non-volatile with which a predetermined program code and data were written in everlastingly. On ROM13, the self diagnostic program (POST) performed to a power up, the code group for hardware input/output operation (BIOS: basic input/output system), etc. are stored.

[0110] The display interface 14 is an exclusive controller for processing actually the drawing instruction which CPU11 publishes, and is supporting the more desirable bit map drawing function equivalent to SVGA (SuperVideo Graphic Array) or XGA (eXtended Graphic Array). Once it is written in a frame buffer (not shown), the screen output of the drawing data processed in the display interface 14 is carried out at a display 21. Indicating equipments 21 are for example, a CRT (Cathode Ray Tube) display, a liquid crystal display display (Liquid Crystal Display), etc.

[0111] The input device interface 15 is equipment for connecting user input devices, such as a keyboard 22 and a mouse 23, to a system 1. The input device interface 15 answers the coordinate directions input through the key input or mouse 23 by the keyboard 22, and generates interruption to CPU11.

[0112] The external instrument interface 17 is equipment for connecting external

devices, such as a hard disk drive (HDD) 24 and the media drive 25, to a system 1. The external instrument interface 17 is based on interface specification, such as IDE (Integrated Drive Electronics) and SCSI (Small Computer System Interface).

[0113] HDD24 is the external storage which carried the magnetic disk as storage support fixed (common knowledge), and excels other external storage in points, such as memory capacity and a data transfer rate. It calls it "install" to the system of a program to place on HDD24 in the condition that a software program can be performed. Usually, the executive program code of CPUs11, such as an operating system, and an application program, a device driver, is stored in HDD24 in un-volatilizing. An example of the application installed in HDD24 is the RIKOMENDO application concerning this example. Moreover, an example of the data file stored in HDD24 is the keyword vector and customer profile which are used for retrieval of information contents.

[0114] Moreover, the media drive 25 is equipment for loading with portable mold media, such as CD (Compact Disc), and MO (Magneto-Optical disc), DVD (Digital Versatile Disc), and accessing a data-logging side. Portable mold media are used in order to mainly move backing up a software program, a data file, etc. as data of a computer-readable format, and these between systems (a sale, a negotiation, and distribution are included). For example, it is possible to distribute data files which distribute the RIKOMENDO application which starts this example through such portable mold media, or are used for retrieval of information contents, such as a keyword vector and a customer profile.

[0115] The network interface 16 can connect a system 1 to LAN (Local Area Network) according to predetermined communications protocols, such as Ethernet. Generally, the network interface 16 is offered with the gestalt of a LAN adapter card, and the PCI bus slot on a mother board (not shown) equips with it, and it is used.

[0116] On LAN, one or more computing systems which work as a server or a client are connected. Moreover, among the computing systems on LAN, a part functions as a router and is connected to the wide area network of the exteriors, such as the Internet.

[0117] On these networks, many computer systems are connected in the transparent condition, and the distributed computing environment is built. It becomes unnecessary for especially each user to recognize the location of a resource object, such as a program and data, under this distributed environment. Moreover, the procedure and the method which are performed in a computer are also dispersedly held and managed on the network. For example, the process which is operating on one certain computer on a network can call and perform procedure of the process which operates on other computers.

[0118] The servers which work on a network may be a file server, a print server, and a database server. Moreover, you may be the WWW (World Wide Web) server which exhibits a HTML (Hyper Text Markup Language) document according to a HTTP

(Hyper Text Transfer Protocol) protocol on the Internet, and the mail server which offers e-mail transceiver service.

[0119] The information retrieval system 10 mentioned above can be positioned on a distributed network as a server which offers RIKOMENDO information to the information contents retrieval demand of books RIKOMENDO from a client etc.

[0120] However, there is not necessarily no need of performing all the treatment processes of an information contents retrieval and RIKOMENDESHON sake on the single computing system 10. For example, after one certain computing system receives an information contents retrieval demand via a WWW server by using the structure of a remote procedure call of CGI (Common Gateway Interface) etc., you may make it leave activation of retrieval processing to other servers on a network.

[0121] Moreover, distribution, such as a software program and data contents, is performed on a network (common knowledge). For example, it is possible to distribute data files, such as a keyword vector for contents retrieval [**** / distributing the RIKOMENDESHON application concerning this example via a network] and a customer profile, via a network.

[0122] The information retrieval system 10 concerning this example is indicating the database used for retrieval of the books as information contents by the list at drawing 9 . However, it is not necessary to accumulate and manage all of these databases on a single computer system, it arranges dispersedly to other one or more database servers and file servers on a network, and you may make it access suitably if needed.

[0123] Subsequently, the information retrieval performed on information retrieval system 10 and generation processing of RIKOMENDO information are explained. Procedure explained below is realized by performing RIKOMENDESHON application concerning this example on the information retrieval system 10 which consists of common computer systems. Moreover, in the following explanation, information retrieval system 10 shall be applied to retrieval of books and RIKOMENDESHON.

[0124] The procedure of the contents retrieval concerning this example is roughly shown in drawing 10 .

[0125] Information retrieval system 10 initializes a temporary table first, in order to search the books as contents (step S1). In order to register the RIKOMENDO information which is the return value of this processing result into the "contents keyword table" for registering the frequency of occurrence (or keyword vector) of a keyword for every "keyword table" for registering the keyword used for contents retrieval, and contents, i.e., books, the "customer profile table" which manages the profile of the customer who is the demand origin of contents retrieval, and a list, a "recommendation contents table" is contained in a temporary table.

[0126] Subsequently, creation processing of a contents keyword is performed (step S2). Although creation processing of a contents keyword is described in more detail to drawing 11 , it is yielded to the after-mentioned for details.

[0127] Subsequently, creation processing of a customer profile is performed (step S3).

Although creation processing of a contents keyword is described in more detail to drawing 12 , it is yielded to the after-mentioned for details.

[0128] Subsequently, selection processing of recommendation contents is performed (step S4). Selection processing of the recommendation contents in this example corresponds to generation of the RIKOMENDO information on books. As mentioned above, order matching type RIKOMENDESHON, and crossover matching mold RIKOMEN day SHON and both hybrid type RIKOMENDESHON are mentioned to books RIKOMENDO. Although order matching type RIKOMENDESHON procedure is described to drawing 13 and the procedure of crossover matching mold RIKOMEN day SHON is described in detail to drawing 14 and drawing 15 , respectively, the after-mentioned is yielded for details.

[0129] In drawing 11 , the detail of the procedure of contents keyword creation is illustrated in the form of a flow chart. Hereafter, it explains according to this flow chart.

[0130] First, it reads one record at a time from a contents table (step S11). And the explanatory note of books is decomposed into a keyword and a part of speech is given (step S12).

[0131] Subsequently, it is confirmed whether the part of speech is contained in the unnecessary part-of-speech list (step S13). Subsequently, it is confirmed whether the keyword is contained in the unnecessary part-of-speech list (step S14).

[0132] The words (for example, vocabulary of the property to appear frequently primarily like an article, "information", and a "system" etc.) which seldom have semantics even if it appears frequently in contents, such as a text document, are contained in the unnecessary part-of-speech list. The part of speech and keyword which are contained in the unnecessary part-of-speech list are excepted from a keyword (step S15).

[0133] Subsequently, it is confirmed whether the record which is in agreement with a contents keyword table (books keyword table) already exists (step S16). When a record in agreement is discovered, only 1 carries out the increment of the frequency of occurrence of this record (step S17). On the other hand, when a record in agreement is not found, this record is added to a contents keyword table (step S18). 1 is given as the frequency of occurrence and zero are given to this new record as weight.

[0134] Processing of steps S13-S18 is repeatedly performed to all keywords (step S19). Moreover, processing of steps S11-S19 is repeatedly performed to all contents (namely, books) (step S20).

[0135] Subsequently, record count is counted and read from a contents keyword table for every keyword (step S21).

[0136] Subsequently, all the read records are added to a keyword table (step S22). The record number of cases is substituted for contents frequency at this time.

[0137] Subsequently, one record is read from a contents keyword table (step S23).

[0138] Subsequently, the record of the same keyword is read from a keyword table (step S24).

[0139] Subsequently, weight (= frequency-of-occurrence /log (total number of contents / contents frequency)) is calculated (step S25).

[0140] Subsequently, the record with which it corresponds in a contents keyword table is updated with the value which the above calculated (step S26).

[0141] And a contents keyword, i.e., a books keyword, is completed by performing processing of steps S23–S26 repeatedly to all contents keywords.

[0142] Moreover, to drawing 12 , the detail of the procedure of customer profile creation is illustrated in the form of a flow chart. Hereafter, it explains according to this flow chart.

[0143] First, it reads one record at a time from the purchasing hysteresis table of the customer who is RIKOMENDO demand origin (step S31).

[0144] And the record whose content ID corresponds is read only one out of a contents keyword table (step S32).

[0145] Subsequently, it is confirmed whether Customer ID, Category ID, and the record whose keyword corresponds exist in a customer profile table (step S33).

[0146] When a record in agreement exists, the weight of the record under this processing is added to the weight of this coincidence record (step S34).

[0147] Subsequently, this record is added to a customer profile table (step S35).

[0148] Processing of the above-mentioned steps S32–S35 is performed to all the records whose content ID corresponds (step S36).

[0149] And processing of steps S31–S36 is performed to all the records in a purchasing hysteresis table (step S37).

[0150] Moreover, to drawing 13 , the detail of order matching type RIKOMENDESHON procedure is illustrated in the form of a flow chart. Hereafter, it explains according to this flow chart.

[0151] First, a keyword vector is read from a customer profile for every customer and category ID (step S41).

[0152] Subsequently, one contents whose category ID corresponds are read from a contents table (step S42).

[0153] Subsequently, the keyword whose content ID corresponds is read from a contents keyword table (step S43).

[0154] And whenever [recommendation] is calculated using the keyword vector by the side of a customer profile, and the keyword vector by the side of contents (step S44). The customer profile is already created according to the procedure shown in drawing 12 .

[0155] Moreover, when the weight of the keyword t_j in Customer's Uk category Ci is set with w'_{jkl} and it sets with the weight w_{ij} of the keyword t_j in Contents (books) d_i , rik_l is expressed by the above-mentioned formula [a-five number] whenever [recommendation / of the contents d_i to Customer's Uk category Ci]. However,

when w'_{jkl} is treated as zero when Keyword t_j does not exist in Customer's U_k category C_l , and Keyword t_j does not exist in Contents d_i , w_{ij} is treated as zero.

[0156] Subsequently, a new record is added to a recommendation contents table (step S45).

[0157] Processing of the above-mentioned steps S42–S45 is repeatedly performed to all the contents, i.e., books, whose category ID corresponds (step S46).

[0158] Moreover, processing of the above-mentioned steps S41–S46 is repeatedly performed to all the keyword vectors in a customer profile (step S47).

[0159] Subsequently, whenever [recommendation] reads a record from a recommendation contents table to high order for every customer (step S48).

[0160] And it is confirmed whether the record corresponding [both sides' / Customer ID and / of content ID] exists in a purchasing hysteresis table (step S49). A record in agreement is deleted from a recommendation contents table (step S54).

[0161] Furthermore, it is confirmed whether the record corresponding [both sides' / Customer ID and / of content ID] exists in a recommendation hysteresis table (step S50). A record in agreement is deleted from a recommendation contents table (step S54).

[0162] Subsequently, recommendation contents are counted for every customer (step S51). And it is confirmed whether the number of contents to recommend the counted number of cases to was exceeded (step S52). If it is less than the number of contents to recommend, additional registration of the record concerned will be carried out at a recommendation hysteresis table (step S53). On the contrary, when the number of contents to recommend has been exceeded, the record concerned is deleted from a recommendation contents table (step S54).

[0163] Processing of the above-mentioned steps S48–S54 is repeatedly performed to all the records in a recommendation contents table (step S55).

[0164] Moreover, to drawing 14 and drawing 15, the detail of the procedure of crossover matching mold RIKOMEN day SHON is illustrated in the form of a flow chart. Hereafter, it explains according to these flow charts.

[0165] First, List A is initialized (step S61).

[0166] And the record which carries out customer ID is read from a purchasing hysteresis table beginning (step S62) to read one record from a customer table (step S63).

[0167] Subsequently, the list of categories ID with purchasing hysteresis is created, and it writes in List A (step S64).

[0168] Subsequently, the record corresponding [Customer's ID] is read for every category ID out of a customer pull file (step S65).

[0169] Subsequently, one contents which fulfill the following conditions 1 are read from a contents table (step S66).

[0170] Conditions 1: Unlike the category ID of a customer profile (a category crosses), Category ID should be contained Category ID in List A.

[0171] Subsequently, the keyword vector whose content ID corresponds is read out of a contents keyword table (step S67). The customer profile is already created according to the procedure shown in drawing 12 .

[0172] Subsequently, whenever [recommendation] is calculated using the keyword vector by the side of a customer profile, and the keyword vector by the side of contents (step S68). The customer profile is already created according to the procedure shown in drawing 12 .

[0173] Moreover, when the weight of the keyword t_j in Customer's Uk category C_i is set with w'_{jkl} and it sets with the weight w_{ij} of the keyword t_j in Contents (books) d_i , rik_l is expressed by the above-mentioned formula [a-five number] whenever [recommendation / of the contents d_i to Customer's Uk category C_i]. However, when w'_{jkl} is treated as zero when Keyword t_j does not exist in Customer's Uk category C_i , and Keyword t_j does not exist in Contents d_i , w_{ij} is treated as zero.

[0174] Subsequently, it is confirmed whether the record which corresponds with Customer ID exists in a recommendation contents table (step S69). When a record in agreement does not exist, additional registration of the record concerned is carried out at a recommendation contents table (step S72).

[0175] Moreover, when the record which corresponds with Customer ID exists in a recommendation contents table, it is confirmed whether it is still higher than whenever [recommendation / which whenever / recommendation / of the record concerned / calculated at step S68] (step S70). In being higher than whenever [recommendation / which was calculated], it updates whenever [on the record concerned / recommendation] with the value of whenever [recommendation / which was calculated] (step S71). Moreover, in not being higher than whenever [recommendation / which was calculated], it skips step S71.

[0176] Category ID performs processing of the above-mentioned steps S66–S72 repeatedly to all the contents corresponding to conditions 1 (step S73).

[0177] Moreover, processing of the above-mentioned steps S65–S73 is repeatedly performed to all the customer profile keyword vectors of the corresponding customer (step S74).

[0178] Moreover, processing of steps S61–S74 is repeatedly performed to all the customers in a customer table (step S75).

[0179] Subsequently, whenever [recommendation] reads one at a time in high order for every customer out of a recommendation contents table (step S76).

[0180] And it is confirmed whether the record corresponding [both sides' / Customer ID and / of content ID] exists in a purchasing hysteresis table (step S77). A record in agreement is deleted from a recommendation contents table (step S82).

[0181] Furthermore, it is confirmed whether the record corresponding [both sides' / Customer ID and / of content ID] exists in a recommendation hysteresis table (step S78). A record in agreement is deleted from a recommendation contents table (step S82).

[0182] Subsequently, recommendation contents are counted for every customer (step S79). And it is confirmed whether the number of contents to recommend the counted number of cases to was exceeded (step S80). If it is less than the number of contents to recommend, additional registration of the record concerned will be carried out at a recommendation hysteresis table (step S81). On the contrary, when the number of contents to recommend has been exceeded, the record concerned is deleted from a recommendation contents table (step S82).

[0183] Processing of the above-mentioned steps S76-S82 is repeatedly performed to all the records in a recommendation contents table (step S83).

[0184] <<comment>>

[1]:G. Salton and C. Buckley, Term-weighting approaches-in-automatic text retrieval., Information Processing-and-Management, 1988, Vol.14, No.5, P.513-523[2]:Marko Balabanovic, and Yoav Shoham, Content-Based, Collaborative Recommendation., Commnications of the ACM, 1997, and Vol. 40, No.3, p66-72 [3] :P aul Resnick and Neophytos Iacovou, Mitesh Suchak and etal., GroupLends:An Open Architecture for Collaborative Filtering of Netnews.Proceedings of the Conference on Computer Supported Cooperative Work and 1994, p175 -186 [4] : NetPerceptionsInc., Recommendation Engine White-Paper.<http://www.netperceptions.com/literature/content/recommendation.pdf>, 2000 [5] : Eric Brill, [http://Rule Based Tagger](http://RuleBasedTagger) and [www.cs.jhu.edu/~brill/ home.html](http://www.cs.jhu.edu/~brill/home.html), 2000 [0185] It has explained in detail about this invention, referring to a specific example more than [addenda]. However, it is obvious that this contractor can accomplish correction and substitution of this example in the range which does not deviate from the summary of this invention.

[0186] At this example, although explained based on the books sale, even if it is the case where the articles and data contents other than retrieval of other goods or goods are discovered, it cannot be overemphasized that this invention can be realized similarly. Moreover, it is not necessarily limited to the contents retrieval on wide area networks, such as the Internet. For example, on the computer system of a stand-alone, even if it is the case where a contents search service is searched on a dialogue screen, it can do [applying this invention similarly or].

[0187] In short, with the gestalt of instantiation, this invention has been indicated and it should not be interpreted restrictively. In order to judge the summary of this invention, the column of the claim indicated at the beginning should be taken into consideration.

[0188]

[Effect of the Invention] As a full account was given above, according to this invention, the outstanding information retrieval system or the outstanding approach of RIKOMENDO [the contents corresponding to a demand of a user / discover and] out of huge information contents can be offered.

[0189] Moreover, according to this invention, based on user profile information, such

as a hobby and taste of a user, and past retrieval hysteresis, purchasing hysteresis, the outstanding information retrieval system or the outstanding approach of desired contents that it can search and RIKOMENDO can be performed can be offered.

[0190] Moreover, according to this invention, when performing business, such as a goods sale, on the Internet, the outstanding information retrieval system or the outstanding approach of RIKOMENDO [the goods which a customer expects / discover and] can be offered.

[0191] Moreover, according to this invention, the outstanding information retrieval system or the outstanding approach of RIKOMENDO [can discover goods with which a customer asks for the unexpected nature other than benefit like a books sale, and / a customer] can be offered.

[0192] Although the information retrieval system and the approach concerning this invention make a contents analysis method the keynote, they can cancel the fault of a contents analysis method by carrying out the device which expands the width of face of RIKOMENDO information. Moreover, also when customer magnitude is small, offer of high quality RIKOMENDO service can be enabled because the number of object contents applies this invention in a very big type of industry like a books sale.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1] It is drawing having shown the procedure in which the information retrieval system concerning this invention generated RIKOMENDO information.

[Drawing 2] It is drawing having shown the books purchasing behavioral model when discovering the books whose interest the customer suited at the actual bookstore assumed in this example.

[Drawing 3] BISAC It is the graph which investigated a customer's number of books purchasing categories using Subject Category.

[Drawing 4] It is drawing having shown the keyword vector creation procedure of a customer profile.

[Drawing 5] It is drawing having shown the creation procedure of the keyword vector for every books.

[Drawing 6] It is drawing having shown order matching type RIKOMENDESHON procedure.

[Drawing 7] It is drawing having shown the procedure of crossover matching mold RIKOMEN day SHON.

[Drawing 8] It is drawing having shown typically the hardware configuration of the information retrieval system 10 with which operation of this invention is presented.

[Drawing 9] The information retrieval system 10 concerning this example is drawing which indicated the database used for retrieval of the books as information contents by the list.

[Drawing 10] It is the flow chart which showed the procedure of the contents retrieval concerning this example roughly.

[Drawing 11] It is the flow chart which showed the procedure of contents keyword creation to the detail.

[Drawing 12] It is the flow chart which showed the procedure of customer profile creation to the detail.

[Drawing 13] It is the flow chart which showed order matching type RIKOMENDESHON procedure to the detail.

[Drawing 14] It is the flow chart which showed the procedure (first half) of crossover matching mold RIKOMEN day SHON to the detail.

[Drawing 15] It is the flow chart which showed the procedure (second half) of crossover matching mold RIKOMEN day SHON to the detail.

[Description of Notations]

10 -- Information retrieval system

11 -- CPU

12 -- RAM

13 -- ROM

14 -- Display interface

15 -- Input device interface

16 -- Network interface

17 -- External instrument interface

19 -- Bus

21 -- Display

22 -- Keyboard

23 -- Mouse

24 -- Hard disk equipment

25 -- Media drive

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号
特開2001-265808
(P2001-265808A)

(43)公開日 平成13年9月28日 (2001.9.28)

(51)Int.Cl. ⁷	識別記号	F I	テーマコード [*] (参考)
G 0 6 F 17/30	3 4 0	G 0 6 F 17/30	3 4 0 A 5 B 0 7 5
	1 1 0		1 1 0 F

審査請求 未請求 請求項の数21 O L (全 24 頁)

(21)出願番号 特願2000-81037(P2000-81037)

(22)出願日 平成12年3月22日(2000.3.22)

(71)出願人 500127690
株式会社スカイソフト
東京都中央区日本橋2丁目7番24号

(72)発明者 廣岡 康雄
東京都文京区大塚3-42-13 コーポ三信
701号

(72)発明者 寺野 隆雄
東京都杉並区和泉3-15-16

(72)発明者 大塚 雄吉
東京都中央区日本橋2丁目7番24号 株式
会社スカイソフト内

(74)代理人 100101801
弁理士 山田 英治

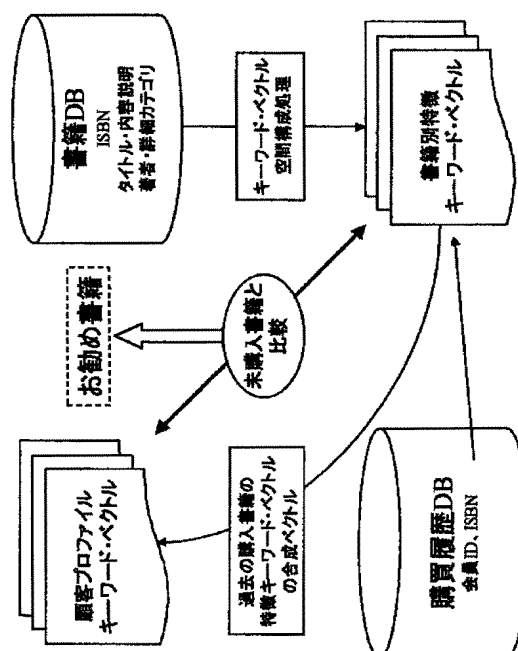
最終頁に続く

(54)【発明の名称】 情報検索システム及び情報検索方法

(57)【要約】

【課題】 有益性と意外性とを兼ね備えた書籍を探し当てて顧客に推薦する。

【解決手段】 顧客が過去に購入した各書籍のキーワード・ベクトルをカテゴリ別に合成して顧客プロフィールを生成する。また、書籍データベースを基にして、各書籍毎にキーワード・ベクトルを生成する。そして、同じカテゴリ間、並びに異なるカテゴリ間の双方の組み合わせで、顧客プロフィールのキーワード・ベクトルと未購入の各書籍のキーワード・ベクトルの類似度を計算することで、有益性と意外性の双方の観点から書籍を検索することができる。類似度の高い未購入書籍をリコメンドする。



【特許請求の範囲】

【請求項1】多数のコンテンツを含む情報空間の中からユーザが望むコンテンツを検索する情報検索システムであって、

前記情報空間上の各コンテンツを所定数のコンテンツ・カテゴリに分類する分類手段と、

前記情報空間上の各コンテンツの特徴ベクトルを生成する特徴ベクトル生成手段と、

ユーザが過去アクセスしたコンテンツの特徴ベクトルを含むユーザ・プロフィールを生成するユーザ・プロフィール生成手段と、

各コンテンツ・カテゴリ毎にユーザ・プロフィールとコンテンツの特徴ベクトルを比較するベクトル比較手段と、

前記比較の結果に基づいて、ユーザが望むコンテンツを特定するコンテンツ特定手段と、を具備することを特徴とする情報検索システム。

【請求項2】前記コンテンツは書籍などのテキスト・ドキュメントを含むコンテンツであり、

前記特徴ベクトル生成手段は、N個のキーワードの各々に対して座標軸を割り当てたN次元キーワード空間上において、コンテンツ中での各キーワードの出現頻度に重み付け因子を掛けた座標値を持つキーワード・ベクトルを生成することを特徴とする請求項1に記載の情報検索システム。

【請求項3】前記ユーザ・プロフィール生成手段は、ユーザが過去アクセスした各コンテンツの特徴ベクトルをコンテンツ・カテゴリ毎に合成して、各コンテンツ・カテゴリ毎の特徴ベクトルからなるユーザ・プロフィールを生成することを特徴とする請求項1に記載の情報検索システム。

【請求項4】前記ユーザ・プロフィール生成手段は、各コンテンツ・カテゴリ毎の特徴ベクトルからなるユーザ・プロフィールを生成し、

前記ベクトル比較手段は、同一カテゴリ内でユーザ・プロフィールの特徴ベクトルとコンテンツの特徴ベクトルを比較することを特徴とする請求項1に記載の情報検索システム。

【請求項5】前記ユーザ・プロフィール生成手段は、各コンテンツ・カテゴリ毎の特徴ベクトルからなるユーザ・プロフィールを生成し、

前記ベクトル比較手段は、異なるカテゴリ間でユーザ・プロフィールの特徴ベクトルとコンテンツの特徴ベクトルを比較することを特徴とする請求項1に記載の情報検索システム。

【請求項6】前記コンテンツ特定手段は、ユーザ・プロフィールと類似度の高いコンテンツをユーザが望むコンテンツとして特定することを特徴とする請求項1に記載の情報検索システム。

【請求項7】前記コンテンツ特定手段は、ユーザ・プロ

ファイルと類似度の高いコンテンツのうちユーザが未アクセスのものをユーザが望むコンテンツとして特定することを特徴とする請求項1に記載の情報検索システム。

【請求項8】多数のコンテンツを含む情報空間の中からユーザが望むコンテンツを検索する情報検索方法であって、

前記情報空間上の各コンテンツの特徴ベクトルを生成する特徴ベクトル生成ステップと、

前記情報空間上の各コンテンツを所定数のコンテンツ・カテゴリに分類する分類ステップと、

ユーザが過去アクセスしたコンテンツの特徴ベクトルを含むユーザ・プロフィールを生成するユーザ・プロフィール生成ステップと、

各コンテンツ・カテゴリ毎にユーザ・プロフィールとコンテンツの特徴ベクトルを比較するベクトル比較ステップと、

前記比較の結果に基づいて、ユーザが望むコンテンツを特定するコンテンツ特定ステップと、

を具備することを特徴とする情報検索方法。

【請求項9】前記コンテンツは書籍などのテキスト・ドキュメントを含むコンテンツであり、

前記特徴ベクトル生成ステップでは、N個のキーワードの各々に対して座標軸を割り当てたN次元キーワード空間上において、コンテンツ中での各キーワードの出現頻度に重み付け因子を掛けた座標値を持つキーワード・ベクトルを生成することを特徴とする請求項8に記載の情報検索方法。

【請求項10】前記ユーザ・プロフィール生成ステップでは、ユーザが過去アクセスした各コンテンツの特徴ベクトルをコンテンツ・カテゴリ毎に合成して、各コンテンツ・カテゴリ毎の特徴ベクトルからなるユーザ・プロフィールを生成することを特徴とする請求項8に記載の情報検索方法。

【請求項11】前記ユーザ・プロフィール生成ステップでは、各コンテンツ・カテゴリ毎の特徴ベクトルからなるユーザ・プロフィールを生成し、前記ベクトル比較ステップでは、同一カテゴリ内でユーザ・プロフィールの特徴ベクトルとコンテンツの特徴ベクトルを比較することを特徴とする請求項8に記載の情報検索方法。

【請求項12】前記ユーザ・プロフィール生成ステップでは、各コンテンツ・カテゴリ毎の特徴ベクトルからなるユーザ・プロフィールを生成し、前記ベクトル比較ステップは、異なるカテゴリ間でユーザ・プロフィールの特徴ベクトルとコンテンツの特徴ベクトルを比較することを特徴とする請求項8に記載の情報検索方法。

【請求項13】前記コンテンツ特定ステップでは、ユーザ・プロフィールと類似度の高いコンテンツをユーザが望むコンテンツとして特定することを特徴とする請求項

8に記載の情報検索方法。

【請求項14】前記コンテンツ特定ステップでは、ユーザ・プロフィールと類似度の高いコンテンツのうちユーザが未アクセスのものをユーザが望むコンテンツとして特定することを特徴とする請求項8に記載の情報検索方法。

【請求項15】多数のコンテンツを含む情報空間の中からユーザが望むコンテンツを検索する情報検索処理をコンピュータ・システム上で実行せしめるためのコンピュータ・ソフトウェアをコンピュータ可読形式で格納するソフトウェア記憶媒体であって、前記コンピュータ・ソフトウェアは、

前記情報空間上の各コンテンツの特徴ベクトルを生成する特徴ベクトル生成ステップと、

前記情報空間上の各コンテンツを所定数のコンテンツ・カテゴリに分類する分類ステップと、

ユーザが過去アクセスしたコンテンツの特徴ベクトルを含むユーザ・プロフィールを生成するユーザ・プロフィール生成ステップと、

各コンテンツ・カテゴリ毎にユーザ・プロフィールとコンテンツの特徴ベクトルを比較するベクトル比較ステップと、

前記比較の結果に基づいて、ユーザが望むコンテンツを特定するコンテンツ特定ステップと、を具備することを特徴とするソフトウェア記憶媒体。

【請求項16】前記コンテンツは書籍などのテキスト・ドキュメントを含むコンテンツであり、

前記特徴ベクトル生成ステップでは、N個のキーワードの各々に対して座標軸を割り当てたN次元キーワード空間において、コンテンツ中での各キーワードの出現頻度に重み付け因子を掛けた座標値を持つキーワード・ベクトルを生成することを特徴とする請求項15に記載のソフトウェア記憶媒体。

【請求項17】前記ユーザ・プロフィール生成ステップでは、ユーザが過去アクセスした各コンテンツの特徴ベクトルをコンテンツ・カテゴリ毎に合成して、各コンテンツ・カテゴリ毎の特徴ベクトルからなるユーザ・プロフィールを生成することを特徴とする請求項15に記載のソフトウェア記憶媒体。

【請求項18】前記ユーザ・プロフィール生成ステップでは、各コンテンツ・カテゴリ毎の特徴ベクトルからなるユーザ・プロフィールを生成し、

前記ベクトル比較ステップでは、同一カテゴリ内でユーザ・プロフィールの特徴ベクトルとコンテンツの特徴ベクトルを比較することを特徴とする請求項15に記載のソフトウェア記憶媒体。

【請求項19】前記ユーザ・プロフィール生成ステップでは、各コンテンツ・カテゴリ毎の特徴ベクトルからなるユーザ・プロフィールを生成し、

前記ベクトル比較ステップは、異なるカテゴリ間でユー

ザ・プロフィールの特徴ベクトルとコンテンツの特徴ベクトルを比較することを特徴とする請求項15に記載のソフトウェア記憶媒体。

【請求項20】前記コンテンツ特定ステップでは、ユーザ・プロフィールと類似度の高いコンテンツをユーザが望むコンテンツとして特定することを特徴とする請求項15に記載のソフトウェア記憶媒体。

【請求項21】前記コンテンツ特定ステップでは、ユーザ・プロフィールと類似度の高いコンテンツのうちユーザが未アクセスのものをユーザが望むコンテンツとして特定することを特徴とする請求項15に記載のソフトウェア記憶媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、膨大な情報コンテンツの中からユーザの要求に合致したコンテンツを探し出すための情報検索システム及び方法に係り、特に、ユーザの趣味・嗜好や過去の探索履歴などのユーザ・プロフィール情報に基づいて所望のコンテンツの探索を行う情報検索システム及び方法に関する。

【0002】更に詳しくは、本発明は、インターネット上で商品販売などのビジネスを行う場合に顧客が望む商品を探し当ててリコメンドするために利用可能な情報検索システム及び方法に係り、特に、書籍販売などのように、顧客が有益性と意外性を兼ね備えた商品を求めるような商品販売ビジネスにおいて適用される情報検索システム及び方法に関する。

【0003】

【従来の技術】昨今における半導体製造その他の工業技術の向上に伴い、パーソナル・コンピュータ（PC）を始めとして比較的低廉な計算システムを市場で提供することが可能となってきた。計算機システムは、企業や研究機関のオフィスや一般家庭に至るまで、広範に普及してきている。さらに、計算機システム同士を相互接続するネットワーク・コンピューティング技術に対する要望も高まってきている。ネットワーク上では、各ユーザ同士で、ファイルやプリンタなどのコンピュータ資源を共有したり、情報の共有・流通・配布・交換などの協働的作業を円滑に行うことができる。

【0004】コンピュータ同士を相互接続するネットワークの形態は様々である。例えば、イーサネット（登録商標）やトークンリングのような局所に敷設されたLAN（Local Area Network）や、LAN同士を専用線などで相互接続して構成されるWAN（Wide Area Network）、さらには、ネットワーク同士の相互接続を繰り返した結果として文字通り世界規模のネットワークへ成長を遂げた「インターネット」（The Internet）など様々である。

【0005】インターネットは、各大学や研究機関等に設置されたサーバが自主的に相互接続を繰り返した結

果、巨大ネットワークへと成長するに至った。インターネット上のサーバ同士は、通常、TCP/IP (Transmission Control Protocol/Internet Protocol) ベースで相互接続されている。

【0006】インターネット上では、WWW (World Wide Web)、News、TELNET (TEletypewriter Network)、FTP (File Transfer Protocol)、Gopherなど、TCP/IPネットワーク上で動作する多数のサービスが公開されている。特に、WWWは、ハイパーリンク構造の情報空間を提供する広域情報検索システムであり、インターネットの爆発的な成長や急速な普及を遂げる最大の要因ともなっている。

【0007】WWWはクライアント・サーバ型で構築される。WWWサーバは、HTTPデーモンと呼ばれるサーバソフトウェアを稼動し、HTML形式で記述された情報資源を格納し、及び／又は、該資源を提供するコンピュータ・システムである。WWWサーバは、HTMLドキュメントを始めとする情報資源を、インターネット上で有償又は無償で公開する。TCP/IPに従えば、これらの情報資源は、URL (Uniform Resource Locator) という形式の識別子によって特定される。HTMLドキュメントは、他の情報資源とのリンク関係を含むことができるが、リンク情報もURL形式で指定される。

【0008】他方のWWWクライアントは、「WWWブラウザ」又は「HTMLブラウザ」と呼ばれるユーザ・エージェントを用いて動作するコンピュータである。米Netscape Communications社の「Netscape Navigator」や、米Microsoft社の「Internet Explorer」がWWWブラウザの代表例である。

【0009】WWWクライアント・ユーザは、ブラウザ画面上の場所ボックスに資源識別子としてのURL文字列を記入することによって、所望の情報資源を指定することができる。WWWブラウザは、通常、HTTP (Hypertext Transfer Protocol) プロトコルに従ってWWWサーバ上の情報資源へのアクセスすなわち情報資源の取り出しを行う。さらに、WWWブラウザは、取得したHTMLドキュメントを解析して、ブラウザ・ウィンドウ画面上で、その内容を表示する（すなわちホームページをブラウザ画面上で組み立てる）。ブラウザ画面に表示されたWWWページ上では、リンク情報は「アンカー」として表示される。ユーザはアンカーをクリックすることで、次の参照先である情報資源を順次指定することができる。すなわち、ユーザは、WWWブラウザ画面を介して、インターネット上でハイパーリンクが張り巡らされたWWW情報資源空間を探索し、多様な情報資源に次々と継ぎ目なくアクセスすることができる（周知）。WWWは、対話入力による簡単な操作で情報検索を行うことができる、先進的なユーザ・インターフェースを備えている。

【0010】インターネットの普及率は既に高く、大学など各種研究機関、企業のオフィス、一般家庭などに深く浸透している。現在、インターネットは、単なる情報配信のみならず、さまざまな商取引の手段として利用されている。いわゆる「インターネット・ビジネス」と呼ばれる商品流通・販売システムである。

【0011】例えば、コンピュータ・ソフトウェアや音楽・映像コンテンツなどのデジタル情報は、インターネット経由で配信することができる。このような場合、CDやDVDなど記録メディアで販売する際に必要とされる商品の流通経路や販売路をほとんど省略することができる。

【0012】また、旧来の通信販売に対してインターネット・ビジネスを適用した場合、1ヶ所のWebサイト上にカタログに相当するデータ・コンテンツをアップロードしておくだけで、商品情報を世界中に瞬間的に配信することができる。この結果、チラシやカタログなどの印刷媒体を配布したり郵送したりする手間やコストを節約することができる。商品情報の即時性があり、商品の売買契約を円滑且つ迅速に成立させることができる。

【0013】インターネット上で商品情報を配信することにより、遠隔地に埋もれた消費者又は商品需要を比較的容易に発掘することができる。商品販売業者側にとっては、商品棚に並べる程の出荷高が見込まれないような少量商品であっても、各地に散在する消費者に対してピンポイントで販売することができる。また、一般消費者側からすれば、輸入品や洋書など最寄の商店には取り扱われておらず割高感のある商品であっても、供給業者を容易に探し出し、低コストで購入することができる。言い換えれば、店舗の敷居が格段と低くなる。

【0014】インターネット・ビジネスは、情報の即時性に優れている。また、他の利点として、商品販売のためのショー・ケースのような物理的な手段をほとんど必要とせず、商品販売及び販売促進に要するイニシャル・コストが低いことが挙げられる。インターネット・ビジネスの発達により、消費者の立場は強まり、商品選択の幅が大きく拡張されたと言われている。さらに、料金の支払いにセキュアな電子決済を採り入れることにより、商品販売のほとんど全工程をネットワーク経由の遠隔操作によって機械化・自動化することができる。

【0015】インターネット・ビジネスでは、基本的には、WWWのようなインタラクティブで先進的なユーザ・インターフェースを介して取引を行うことができる。また、インターネット・ビジネスを展開する企業では、顧客とインタラクティブに結ばれているという点を活かして、顧客毎にコンテンツやサービスをカスタマイズするパーソナライズ化が志向されている。

【0016】しかしながら、顧客毎に個別の対応を行うことは、WWWや電子メールにおける既存の技術で実現可能である一方、顧客毎に適切なコンテンツを準備した

り推薦することは容易ではない。このような課題に対しては、自動的に推薦情報を生成する「リコメンダー・システム」が期待されている。

【0017】例えば、インターネット上で洋書などの書籍販売を行うようなビジネスに対してリコメンダー・システムを適用することにより、各読者に対して好みに応じた新刊書を提供することができ、顧客サービスを拡充するとともに、売上の増大にもつながる。

【0018】リコメンダー・システムには、従来より、「コンテンツ分析方式」及び「協調フィルタリング方式」と呼ばれる2方式が存在している。

【0019】コンテンツ分析方式とは、リコmend対象となるコンテンツの特徴表現とユーザ興味表現であるユーザ・プロフィールの比較を行うことにより、ユーザの興味により近いコンテンツを選択して、リコmendする方法である。洋書販売に適用する場合、コンテンツの特徴表現は、各書籍のタイトル、著者名、ジャンル、キーワードなどからなる書籍データベースとして蓄積される。また、ユーザ・プロフィールは、ユーザの趣味や特技などのプロフィールの他、過去の書籍購買履歴や探索履歴で構成される。

【0020】コンテンツ分析のために、情報検索分野の周知技術を適用することができる。例えば、重み付きタームによるベクトル空間モデル^[1]型の検索のような、コンテンツに関する特徴ベクトル同士を比較するような分析手法を利用することができる。すなわち、N個のキーワードの各々に対して座標軸を割り当てたN次元空間において、コンテンツ中での各キーワードの出現回数に重み付け因子を掛けた座標値を持つベクトルすなわち「キーワード・ベクトル」を求め、キーワード・ベクトル同士の類似度を基にして所望のコンテンツを検索するものである。類似度は、例えば、ベクトルの内積や距離に基づいて求めることができる。

【0021】しかしながら、コンテンツ分析方式の欠点として、次に挙げる2点が指摘されている^[2]。

(1) 検索対象とするコンテンツが、コンテンツ分析技術に依存する。現状では、情報検索技術が適用できるテキスト・ドキュメントに制限される。

【0022】(2) ユーザが高く評価した情報に類似した情報ばかりが抽出される傾向がある。

【0023】特に、情報検索を洋書などの書籍のリコメンデーションに利用する場合、以前に購入した書籍と類似した内容のものばかりが常に好まれる訳ではない。そもそも読書は知的欲求に基づく行為であるから、ユーザ(読者)の趣味・嗜好に合致するなど有益性があるだけでなく、意外性や新規性のある書籍をリコmendすることが好ましい。類似内容の書籍ばかりリコメンデーションを繰り返すと、顧客はやがて飽きてしまう。したがって、コンテンツ分析方式が持つ上記(2)の欠点は、書籍販売ビジネスに適用する上で致命的になりかねない。

い。

【0024】また、協調フィルタリング方式は、コンテンツ分析を行わずに、複数のユーザ間でのコンテンツ評価情報を分析することによってリコmend情報を生成する方式である。まず、ユーザ間のコンテンツ評価情報を比較することにより、類似した興味を持つ仲間をグループ化する。そして、自分の仲間が高く評価したコンテンツのうち、自分が未評価のものがリコmendされる。協調フィルタリング方式は、NetPerceptions^[3]^[4]に代表される、いくつかの商用ソリューションにおいて既に採用されている。

【0025】協調フィルタリング方式は、コンテンツ分析を行わないために、対象とするコンテンツに制限がないという利点があり、テキスト・ドキュメント以外のほとんどすべてのメディアに対しても適用可能である。一方で、協調フィルタリング方式の欠点としては、次に挙げる4点が指摘されている^[2]。

【0026】(1) 対象コンテンツの量に対して、十分なユーザ数と評価情報数が必要になる。ユーザ数が不足する場合には、カバレッジが狭くなり、すべてのコンテンツのうち限られた一部しかリコmend対象とすることができなくなる。また、特異な嗜好を持つユーザには仲間が見つからない結果として、リコmend自体が不能になるなどが懸念される。

【0027】(2) 新しいコンテンツが登録されても、誰かが評価するまではリコmend対象にならない。

【0028】(3) コンテンツ分析を行わないため、内容が類似しているがIDが異なるコンテンツの評価が高いユーザ同士を仲間にすることができない。

【0029】(4) コンテンツ評価の誤入力、他ユーザへのリコmendに与える影響を排除することが困難である。

【0030】書籍を対象にリコmend・サービスを考察した場合、基本的には、コンテンツ方式と協調フィルタリング方式のどちらも適用可能である。但し、上述したように各方式にはそれぞれ欠点があることから、次の点が懸念される。

【0031】(1) コンテンツ分析方式では、リコmendされる書籍が、顧客が過去に購入した書籍に類似したものばかりになる可能性が高い。類似内容の書籍ばかりリコメンデーションを繰り返すと、顧客はやがて飽きてしまう。

【0032】(2) 協調フィルタリング方式では、書籍販売では対象とするコンテンツ数が膨大であるため、リコmend・サービスの対象となる顧客数も相当な規模でなければ、質の高いサービスが期待できない。

【0033】

【発明が解決しようとする課題】本発明の目的は、膨大な情報コンテンツの中からユーザの要求に合致したコンテンツを探し出してリコmendすることができる、優れ

た情報検索システム及び方法を提供することにある。

【0034】本発明の更なる目的は、ユーザの趣味・嗜好や過去の探索履歴・購買履歴などのユーザプロフィール情報に基づいて所望のコンテンツの探索しりこめンドを行うことができる、優れた情報検索システム及び方法を提供することにある。

【0035】本発明の更なる目的は、インターネット上で商品販売などのビジネスを行う場合において、顧客が期待する商品を探し当ててリコメンドすることができる、優れた情報検索システム及び方法を提供することにある。

【0036】本発明の更なる目的は、書籍販売などのように、顧客が有益性の他に意外性を求めるような商品を探し当てリコメンドすることができる、優れた情報検索システム及び方法を提供することにある。

【0037】

【課題を解決するための手段】本発明は、上記課題を参酌してなされたものであり、その第1の側面は、多数のコンテンツを含む情報空間の中からユーザが望むコンテンツを検索する情報検索システム又は方法であって、前記情報空間上の各コンテンツを所定数のコンテンツ・カテゴリに分類する分類手段又はステップと、前記情報空間上の各コンテンツの特徴ベクトルを生成する特徴ベクトル生成手段又はステップと、前記情報空間上の各コンテンツを所定数のコンテンツ・カテゴリに分類する分類手段又はステップと、ユーザが過去アクセスしたコンテンツの特徴ベクトルを含むユーザ・プロフィールを生成するユーザ・プロフィール生成手段又はステップと、各コンテンツ・カテゴリ毎にユーザ・プロフィールとコンテンツの特徴ベクトルを比較するベクトル比較手段又はステップと、前記比較の結果に基づいて、ユーザが望むコンテンツを特定するコンテンツ特定手段又はステップと、を具備することを特徴とする情報検索システム又は方法である。

【0038】但し、本発明に係る情報検索システム又は方法において、書籍のようにあらかじめコンテンツ・カテゴリが規定されている商品を検索対象コンテンツとして取り扱う場合には、上記の分類手段又はステップを省略して、既存のコンテンツ・カテゴリをそのまま利用してもよい。書籍カテゴリの一例は、BISAC (Book Industry Systems Advisory Committee) が提供する“BISAC Subject Category”である。

【0039】本発明の第1の側面に係る情報検索システム又は方法において、前記コンテンツは書籍などのテキスト・ドキュメントを含むコンテンツであってもよい。このような場合、前記特徴ベクトル生成手段又はステップは、N個のキーワードの各々に対して座標軸を割り当てたN次元キーワード空間上において、コンテンツ中での各キーワードの出現頻度に重み付け因子を掛けた座標値を持つキーワード・ベクトルを生成するようにしても

よい。

【0040】また、前記ユーザ・プロフィール生成手段又はステップは、ユーザが過去アクセスした各コンテンツの特徴ベクトルをコンテンツ・カテゴリ毎に合成して、各コンテンツ・カテゴリ毎の特徴ベクトルからなるユーザ・プロフィールを生成するようにしてもよい。

【0041】また、前記ユーザ・プロフィール生成手段又はステップにおいて、各コンテンツ・カテゴリ毎の特徴ベクトルからなるユーザ・プロフィールを生成するような場合には、前記ベクトル比較手段又はステップは、同一カテゴリ内でユーザ・プロフィールの特徴ベクトルとコンテンツの特徴ベクトルを比較する（順マッチング）ようにしてもよい。あるいは、前記ベクトル比較手段又はステップは、異なるカテゴリ間でユーザ・プロフィールの特徴ベクトルとコンテンツの特徴ベクトルを比較する（交差マッチング）ようにしてもよい。

【0042】また、前記コンテンツ特定手段又はステップは、ユーザ・プロフィールと類似度の高いコンテンツをユーザが望むコンテンツとして特定して、ユーザにリコメンドするようにしてもよい。

【0043】また、前記コンテンツ特定手段又はステップは、ユーザ・プロフィールと類似度の高いコンテンツのうちユーザが未アクセスのものをユーザが望むコンテンツとして特定して、ユーザにリコメンドするようにしてもよい。

【0044】また、本発明の第2の側面は、多数のコンテンツを含む情報空間の中からユーザが望むコンテンツを検索する情報検索処理をコンピュータ・システム上で実行せしめるためのコンピュータ・ソフトウェアをコンピュータ可読形式で格納するソフトウェア記憶媒体であって、前記コンピュータ・ソフトウェアは、前記情報空間上の各コンテンツの特徴ベクトルを生成する特徴ベクトル生成ステップと、前記情報空間上の各コンテンツを所定数のコンテンツ・カテゴリに分類する分類ステップと、ユーザが過去アクセスしたコンテンツの特徴ベクトルを含むユーザ・プロフィールを生成するユーザ・プロフィール生成ステップと、各コンテンツ・カテゴリ毎にユーザ・プロフィールとコンテンツの特徴ベクトルを比較するベクトル比較ステップと、前記比較の結果に基づいて、ユーザが望むコンテンツを特定するコンテンツ特定ステップと、を具備することを特徴とするソフトウェア記憶媒体である。

【0045】

【作用】本発明に係る情報検索システム及び方法は、書籍等のテキスト・ドキュメントからなるコンテンツの探索やリコメンデーションなどに好適に適用することができる。

【0046】本発明に係る情報検索システム及び方法では、顧客が過去に購入した各書籍の特徴ベクトルをカテゴリ別に合成して、カテゴリ毎の特徴ベクトルからなる

顧客プロフィールを生成する。また、書籍データベースを基にして、各書籍に対してキーワード・ベクトル空間構成処理を実行することで、各書籍毎に特徴ベクトルを生成する。

【0047】そして、顧客プロフィールの特徴ベクトルと未購入の各書籍が持つ特徴ベクトルの類似度を計算して、類似度の高い未購入書籍をリコメンドする。同じカテゴリ間（順マッチング）、並びに異なるカテゴリ間（交差マッチング）の各々の組み合わせで特徴ベクトルの類似度計算を行うことで、有益性と意外性の双方の観点から書籍を検索することができる。

【0048】本発明において取り扱う特徴ベクトルは、例えば、テキスト・ドキュメントなどのコンテンツ中に散在するキーワードの出現頻度を表現した「キーワード・ベクトル」でよい。キーワード・ベクトルは、N個のキーワードの各々に対して座標軸を割り当てたN次元キーワード空間上において、コンテンツ中での各キーワードの出現回数に重み付け因子を掛けた座標値を持つベクトルのことである。

【0049】本発明に係る情報検索システム及び方法は、コンテンツ分析方式を基調とするが、リコメンド情報の幅を広げる工夫をすることにより、コンテンツ分析方式の欠点を解消することができる。また、書籍販売のように対象コンテンツ数が非常に大きな業種において本発明を適用することで、顧客規模が小さい場合にも、質の高いリコメンド・サービスを提供可能にすることができる。

【0050】本発明の第2の側面に係るソフトウェア記憶媒体は、例えば、様々なプログラム・コードを実行可能な汎用コンピュータ・システムに対して、コンピュータ・ソフトウェアをコンピュータ可読な形式で提供する媒体である。このような媒体は、例えば、CD（Compact Disc）やFD（Floppy Disc）、MO（Magnet-Optical disc）などの着脱自在で可搬性の記憶媒体である。

【0051】このようなソフトウェア記憶媒体は、コンピュータ・システム上で所定のコンピュータ・ソフトウェアの機能を実現するための、コンピュータ・ソフトウェアと記憶媒体との構造上又は機能上の協働的関係を定義したものである。換言すれば、本発明の第2の側面に係るソフトウェア記憶媒体を介して所定のコンピュータ・ソフトウェアをコンピュータ・システムにインストールすることによって、コンピュータ・システム上では協働的作用が発揮され、本発明の第1の側面に係る情報検索システム又は方法と同様の作用効果を得ることができる。

【0052】本発明のさらに他の目的、特徴や利点は、後述する本発明の実施例や添付する図面に基づくより詳細な説明によって明らかになるであろう。

【0053】

【発明の実施の形態】本発明に係る情報検索システム

は、例えば書籍販売に適用して、膨大な数の書籍の山から各顧客が望む書籍を探し出してリコメンドすなわち購買を推薦することができる。

【0054】図1には、本発明に係る情報検索システムがリコメンド情報を生成する手順を図解している。

【0055】テキスト・ドキュメントを対象としたコンテンツ分析型リコメンデーションでは、顧客興味の実現である顧客プロフィールやコンテンツの特徴表現としてキーワード・ベクトルを用いるのが一般的であるが、本発明に係る情報検索システムの場合も同様である。キーワード・ベクトルとは、N個のキーワードの各々に対して座標軸を割り当てたN次元空間上において、コンテンツ中での各キーワードの出現回数に重み付け因子を掛けた座標値を持つN次元ベクトルのことである（前述）。

【0056】本情報検索システムでは、書籍情報（書籍データベース）と顧客の購買履歴情報（購買履歴データベース）とを用いて、顧客プロフィールのキーワード・ベクトルを生成する。ここでは、顧客プロフィールのキーワード・ベクトルは、顧客が過去に購入した各書籍のキーワード・ベクトルをカテゴリ別に合成して生成されるベクトルで構成される。購買履歴情報は、例えば、顧客の会員識別子（ID）や、過去に購入した書籍の識別子（ISBN：International Standard BookNumber）などで構成される。

【0057】他方、書籍データベースを基にして、各書籍に対してキーワード・ベクトル空間構成処理を実行することで、各書籍毎にキーワード・ベクトルを生成することができる。書籍データベースは、各書籍のタイトル、内容説明（要約／抄録など）、著者名、カテゴリなどで構成される。

【0058】そして、顧客プロフィールのキーワード・ベクトルと未購入の各書籍のキーワード・ベクトルの類似度を計算する。類似度の高い未購入書籍を顧客に対してリコメンドすることができる。

【0059】次いで、顧客プロフィールの作成処理について詳解する。

【0060】本発明では、顧客が実際の書店で興味にあった書籍を探し出す行動は、図2に示すような書籍購買行動モデルに従うことを仮定する。図示の通り、購入書籍の決定に至るまでの行動は、興味分野にあった書棚を選択する行動と、該当する書棚の書籍の背表紙や目次の情報を利用して興味に沿った書籍を選択する行動の2つに大別される。それぞれの行動の特徴から、前者を「分類型選択行動」と呼び、後者を「検索型選択行動」と呼ぶことができる。

【0061】本発明に係る情報検索システムでは、図2に示す書籍購買行動モデルに基づき、分類型選択行動に相当する書籍カテゴリと、検索型選択行動の基準となるキーワード・ベクトルという、2種類の顧客興味表現系を採用する。

【0062】書籍カテゴリは、本発明に係る情報検索システムのためにゼロから構築してもよいし、既存の書籍カテゴリを適用することも可能である。既存の書籍カテゴリとしては、例えば、BISAC (Book Industry Systems Advisory Committee) が提供する“BISAC Subject Category”を挙げることができる。BISAC Subject Category では、Art、Business & Economics、Cooking、Computersなどの49種類のカテゴリに分類している。

【0063】図3には、BISAC Subject Categoryを用いて、顧客の書籍購買カテゴリ数を調べたグラフを示している。複数の書籍を購入した人ほど複数のカテゴリにまたがって書籍を購入している場合が多く、一般書籍に対する興味の幅は非常に広い、という傾向があることを理解できるであろう。

【0064】WWW (World Wide Web) を始めとする各種の情報検索システムが充実したインターネット上で書籍販売ビジネスを行う場合、ユーザすなわち顧客が自ら販売サイトにアクセスして、目的の書籍を検索するのが一般的である。検索には、書籍タイトル、著者名、出版社やキーワードなどが用いられる。複数のカテゴリに興味を持っている顧客の場合、検索する書籍カテゴリ毎にまったく異なるキーワードを使い分けるであろうことは、容易に想像できる。

【0065】したがって、顧客興味を表現する顧客プロフィールを作成する場合にも、顧客の興味の幅が広い場合には、複数のキーワード・ベクトルを作成する必要があると思料される。

【0066】本発明に係る情報検索システムでは、書籍カテゴリを顧客興味の単位と仮定して、購買書籍カテゴリ別の複数のキーワード・ベクトルにより顧客プロフィールを構成するようにした。

【0067】具体的な顧客プロフィールとしては、キーワード・ベクトルを用いる。キーワード・ベクトルとは、N個のキーワードの各々に対して座標軸を割り当てたN次元キーワード空間上において、コンテンツ中での各キーワードの出現回数に重み付け因子を掛けた座標値を持つベクトルのことである(前述)。

【0068】書籍データベース上の項目から、書籍の特徴を代表する項目として、タイトル、内容説明、著者、サブカテゴリ(49分類をさらに詳細に約3,700に分類したもの)を用いて、キーワード・ベクトルを構成する。一般には、タイトルや内容説明といったテキスト形式のデータ中の語句をキーワードとして利用するが、著者やサブカテゴリの情報も1つのキーワードとしてベクトル中に含めるようにしてもよい。

【0069】本実施例では、顧客プロフィールは、書籍カテゴリ毎のキーワード・ベクトルで構成される。図4には、顧客プロフィールのキーワード・ベクトル作成処理手順を図解している。まず、書籍毎にキーワード・ベクトルを付与する。次いで、この顧客が過去に購買した

各書籍のキーワード・ベクトルを、書籍カテゴリ毎に合算することによって、顧客の購買書籍カテゴリ毎のキーワード・ベクトルを作成する。

【0070】また、図5には、書籍毎のキーワード・ベクトルの作成手順を図解している。但し、書籍のタイトルと内容説明についてはテキスト処理を行う必要がある。テキスト処理の内容は、次の手順である。

【0071】(1) テキストに含まれる各単語に品詞を付与する。品詞を付与するためのツールとしては、Brill's English tagger [5]を用いる。

【0072】(2) 不要語の削除は、品詞情報を用いて名詞・形容詞・外来語以外の品詞を削除する。その後、事前に作成した不要語リストを適用する。

【0073】(3) 削除されずに残った単語を単語キーワードとして利用する。

【0074】(4) 複合語キーワードは、文章中に名詞が連続して出現した場合に、その組み合わせを複合語キーワードとして利用する。

【0075】著者、書籍カテゴリ、サブカテゴリについては、前述の単語キーワードや複合語キーワードと同様に重み付けを行い、キーワード・ベクトルに加える。本実施例では、重み付けの方法は、語句の出現頻度TF (Term Frequency) と逆文献頻度IDF (Inverse Document Frequency) を用いた、TFIDF [1]と呼ばれる重み付け手法を用いる。

【0076】以下の説明では、i番目の書籍d_iにおける単語・複合語・著者・サブカテゴリのj番目のキーワードt_jの出現頻度をf_{ij}とし、また、全書籍数Nの中でキーワードt_jが出現する書籍数をn_jとおくこととする。また、キーワードの種別毎に設定した重みをkとする。本明細書中では、キーワード種別毎の重みは、経験的にサブカテゴリについてはk=2とし、それ以外についてはk=1とした。

【0077】また、同一のキーワードが過剰に出現した場合の影響を緩和するために、i番目の書籍d_iにおけるj番目のキーワードt_jの出現頻度f_{ij}をそのまま用いるのではなく、対数表現を採り入れた以下の式で定義されるTF_{ij}を用いて各キーワードの出現頻度を評価することとした。

【0078】

【数1】

$$TF_{ij} = \log_2(f_{ij} \cdot k + 1)$$

【0079】また、逆文献頻度に関しては、対数表現を採り入れて、以下の式で定義されるIDF_jを用いることとした。

【0080】

【数2】

$$IDF_j = \log_2(N/n_j)$$

【0081】また、書籍 d_j におけるキーワード t_j に付与される重み w_{ij} は、上記の各式〔数1〕及び〔数2〕を用いて次式となる。

$$w_{ij} = TF_{ij} \times IDF_j = \log_2(f_{ij} \cdot k + 1) \cdot \log_2(N/n_j)$$

【0083】最終的な書籍毎のキーワード・ベクトルは、キーワード t_j の重み w_j の降順に上位12個のキーワードを各書籍に付与する。但し、本発明を実現する上で、キーワードの上限を設定する必要は必ずしもなく、また、上限値が12個に限定されるものでもない。

【0084】図4に示すように、顧客プロフィールは、書籍カテゴリ毎の購買書籍のキーワード・ベクトルの和、すなわちカテゴリ毎の合成ベクトルで構成される。顧客 U_k のカテゴリ C_l についてのキーワード t_j の重み w'_{jkl} は、下式のように定義される。

【0085】

〔数4〕

$$w'_{jkl} = \sum_i (w_{ij} \cdot b_{ik} \cdot s_{il})$$

【0086】但し、上式〔数4〕において、 b_{ik} は、顧客 U_k が書籍 d_i を購入していた場合に1とし、それ以外は0とする。また、 s_{il} は、書籍 d_i がカテゴリ C_l に属している場合に1とし、それ以外は0とする。

【0087】顧客にリコメンドする書籍の選択は、顧客プロフィールのキーワード・ベクトルと書籍の特徴を表すキーワード・ベクトルの類似度を計算することによって行う。類似度の計算方法には、キーワード・ベクトル間のなす角のコサインを用いる。

【0088】顧客 U_k のカテゴリ C_l におけるキーワード t_j の重みを w'_{jkl} とおき、また、書籍 d_i におけるキーワード t_j の重み w_{ij} とおくと、顧客 U_k のカテゴリ C_l に対する書籍 d_i の推薦度 r_{ikl} は、次式の通りとなる。

【0089】

〔数5〕

$$r_{ikl} = \frac{\sum_j w_{ij} w'_{jkl}}{\sqrt{\sum_j w_{ij}^2} \sqrt{\sum_j w'_{jkl}^2}} \quad (0 \leq r_{ikl} \leq 1)$$

【0090】次いで、書籍のリコメンデーション処理について説明する。

【0091】キーワード・ベクトルのような特徴ベクトルを用いたコンテンツ分析手法では、一般に、ベクトル同士の類似度に基づいて特徴比較が行われる。

【0092】本実施例では、カテゴリ毎に特徴比較を行う。より具体的には、同一の書籍カテゴリに含まれるキーワード・ベクトル同士を比較する「順マッチング型リコメンデーション」と、異なる書籍カテゴリに含まれるキーワード・ベクトル同士を比較する「交差マッチング型リコメンデーション」の2通りを行う。

【0093】図6には、順マッチング型リコメンデーション

【0082】

〔数3〕

の処理手順を図解している。

【0094】顧客の興味表現すなわち顧客プロフィールは、過去の購買履歴のある各書籍カテゴリにおけるキーワード・ベクトルの合成ベクトルで構成される。図6に示す例では、顧客が過去に購入したことがある書籍カテゴリA及びBの各々において、顧客プロフィールを表現するキーワード・ベクトルが生成される。

【0095】他方、書籍データベースを基に、書籍毎のキーワード・ベクトルが生成されており（図4を参照のこと）、このうち、顧客の購買履歴が残された書籍カテゴリA及びBの各々に含まれる書籍のキーワード・ベクトルを抽出する。

【0096】さらに、同じ書籍カテゴリA又はB内で、顧客プロフィールのキーワード・ベクトルと書籍のキーワード・ベクトルの比較すなわち類似度を計算する。類似度の計算には、上記の〔数5〕を用いればよい。そして、類似度の高い書籍のうち、顧客が未購入のものを検索結果として顧客にリコメンドする。

【0097】順マッチング型リコメンデーションでは、まったく興味のないカテゴリの書籍が、いくつかのキーワードが一致したというだけでリコメンドされてしまう、すなわち不要又は無益なリコメンデーションを防ぐことができる。

【0098】また、図7には、交差マッチング型リコメンデーションの処理手順を図解している。

【0099】顧客の興味表現すなわち顧客プロフィールは、過去の購買履歴のある各書籍カテゴリにおけるキーワード・ベクトルの合成ベクトルで構成される。図7に示す例では、顧客が過去に購入したことがある書籍カテゴリA及びBの各々において、顧客プロフィールを表現するキーワード・ベクトルが生成される。また、書籍データベースを基に、顧客の購買履歴が残された書籍カテゴリA及びBの各々に含まれる書籍のキーワード・ベクトルを抽出する（同上）。

【0100】さらに、異なる書籍カテゴリの間で、顧客プロフィールのキーワード・ベクトルと書籍のキーワード・ベクトルの比較を行う。図7に示す例では、書籍カテゴリAの顧客プロフィールと書籍カテゴリBの書籍キーワード・ベクトルとを比較するとともに、書籍カテゴリBの顧客プロフィールと書籍カテゴリAの書籍キーワード・ベクトルとを比較する。ベクトルの比較すなわち類似度の計算には、上記の〔数5〕を用いればよい。そして、類似度の高い書籍のうち、顧客が未購入のものを検索結果として顧客にリコメンドする。

【0101】交差マッチング型リコメンデーションによ

れば、キーワード・ベクトルと書籍特徴の書籍カテゴリの関係を変差させることにより、意外性の高いリコメンド情報を生成することができる。

【0102】例えば、図2に示した書籍購買行動モデルの書籍探索フェーズのように、目的の書棚から必要なキーワードを持つ書籍を探す行動を行っている際に、普段から興味を持っているがまったく別の分野のキーワードを含む書籍を偶然見つけるようなケースである。交差マッチング型リコメンデーションは、意図的にこのような効果を狙う手法である。

【0103】分かり易い例としては、Science FictionのカテゴリからStar Trekに関する書籍とCookingカテゴリからいくつかの書籍を購入している顧客に対するリコメンデーションを考察してみる。交差マッチング型リコメンデーションによれば、Cookingカテゴリの書籍をキーワード“Star Trek”を用いて検索し、“Star Trek Cookbook”といった書籍を紹介することができる。

【0104】但し、交差マッチングにおいては、比較の対象となる書籍カテゴリを、過去に購入実績があり顧客が興味を持っている書籍カテゴリに限定することが、リコメンデーションの精度を向上する上で好ましいと思料する。

【0105】

【実施例】以下、図面を参照しながら本発明の実施例を詳解する。

【0106】図8には、本発明の実施に供される情報検索システム10のハードウェア構成を模式的に示している。情報検索システム10は、専用のハードウェア装置としてデザインしてもよいが、一般的な計算機システム上で所定のアプリケーション（本明細書中では、仮に「リコメンド・アプリケーション」とする）を実行することによっても実現される。計算機システムの一例は、米IBM社のPC/AT互換機又はその後継機である。以下、システム10の各部について説明する。

【0107】システム10のメイン・コントローラであるCPU（Central Processing Unit）11は、オペレーティング・システム（OS）の制御下で、各種のアプリケーションを実行するようになっている。OSは、例えばUNIX（登録商標）、あるいは米マイクロソフト社のWindows 98/NTでよい。

【0108】図示の通り、CPU11は、バス19によって他の機器類と相互接続されている。バス19上の各機器にはそれぞれ固有のメモリ・アドレス又はI/Oアドレスが付与されており、CPU11はこれらアドレスによって機器アクセスが可能となっている。バス19の一例はPCI（Peripheral Component Interconnect）バスである。

【0109】RAM（Random Access Memory）12は、CPU11において実行されるプログラム・コードをロードしたり、実行中の作業データを一時保管するために

使用される揮発性の記憶装置であり、通常は、複数のDRAM（dynamic RAM）チップで構成される。また、ROM（Read Only Memory）13は、所定のプログラム・コードやデータが恒久的に書き込まれた不揮発性の記憶装置である。ROM13上には、例えば、電源投入時に実行する自己診断プログラム（POST）や、ハードウェア入出力操作のコード群（BIOS：基本入出力システム）などが格納されている。

【0110】ディスプレイ・インターフェース14は、CPU11が発行する描画命令を実際に処理するための専用コントローラであり、より好ましくはSVGA（SuperVideo Graphic Array）又はXGA（eXtended Graphic Array）相当のビットマップ描画機能をサポートしている。ディスプレイ・インターフェース14において処理された描画データは、例えばフレーム・バッファ（図示しない）に一旦書き込まれた後、表示装置21に画面出力される。表示装置21は、例えば、CRT（Cathode Ray Tube）ディスプレイや、液晶表示ディスプレイ（Liquid Crystal Display）などである。

【0111】入力機器インターフェース15は、キーボード22やマウス23などのユーザ入力機器をシステム1に接続するための装置である。入力機器インターフェース15は、キーボード22によるキー入力又はマウス23を介した座標指示入力に応答して、CPU11に対して割り込みを発生する。

【0112】外部機器インターフェース17は、ハード・ディスク・ドライブ（HDD）24やメディア・ドライブ25などの外部装置をシステム1に接続するための装置である。外部機器インターフェース17は、例えば、IDE（Integrated Drive Electronics）やSCSI（Small Computer System Interface）などのインターフェース規格に準拠する。

【0113】HDD24は、記憶担体としての磁気ディスクを固定的に搭載した外部記憶装置であり（周知）、記憶容量やデータ転送速度などの点で他の外部記憶装置よりも優れている。ソフトウェア・プログラムを実行可能な状態でHDD24上に置くことをプログラムのシステムへの「インストール」と呼ぶ。通常、HDD24には、オペレーティング・システムや、アプリケーション・プログラム、デバイス・ドライバなどのCPU11の実行プログラム・コードが不揮発的に格納されている。HDD24にインストールされるアプリケーションの一例は、本実施例に係るリコメンド・アプリケーションである。また、HDD24に格納されるデータ・ファイルの一例は、情報コンテンツの検索に利用されるキーワード・ベクトルや顧客プロファイルである。

【0114】また、メディア・ドライブ25は、CD（Compact Disc）やMO（Magneto-Optical disc）、DVD（Digital Versatile Disc）などの可搬型メディアを装填して、データ記録面にアクセスするための装置であ

る。可搬型メディアは、主として、ソフトウェア・プログラムやデータ・ファイルなどをコンピュータ可読形式のデータとしてバックアップすることや、これらをシステム間で移動（販売・流通・配布を含む）する目的で使用される。例えば、このような可搬型メディアを媒介にして、本実施例に係るリコmend・アプリケーションを配布したり、情報コンテンツの検索に利用されるキーワード・ベクトルや顧客プロファイルなどのデータ・ファイルを配布することが可能である。

【0115】ネットワーク・インターフェース16は、Ethernetなどの所定の通信プロトコルに従って、システム1をLAN（Local Area Network）に接続することができる。ネットワーク・インターフェース16は、一般に、LANアダプタ・カードの形態で提供され、マザーボード（図示しない）上のPCIバス・スロットの装着して用いられる。

【0116】LAN上には、サーバ又はクライアントとして稼動する1以上の計算機システムが接続されている。また、LAN上の計算機システムのうち一部はルータとして機能し、インターネットなどの外部の広域ネットワークに接続されている。

【0117】これらネットワーク上では、多数の計算機システムがトランスペアレントな状態で接続され、分散コンピューティング環境が構築されている。かかる分散環境下では、各ユーザはプログラムやデータなど資源オブジェクトの所在を特に認識する必要がなくなる。また、コンピュータにおいて実行される手続きやメソッドも、ネットワーク上で分散して保持され、管理されている。例えば、ネットワーク上のある1つのコンピュータ上で動作しているプロセスが、他のコンピュータ上で動作するプロセスの手続きを呼び出して実行させることができる。

【0118】ネットワーク上で稼動するサーバは、例えば、ファイル・サーバ、プリント・サーバ、データベース・サーバであってもよい。また、インターネット上でHTTP（Hyper Text Transfer Protocol）プロトコルに従ってHTML（Hyper Text Markup Language）ドキュメントを公開するWWW（World Wide Web）サーバや、メール送受信サービスを提供するメール・サーバであってもよい。

【0119】上述した情報検索システム10は、分散ネットワーク上において、クライアントからの書籍リコmendなどの情報コンテンツ検索要求に対してリコmend情報を提供するサーバとして位置付けることができる。

【0120】但し、情報コンテンツ検索及びリコmendーションのための処理プロセスをすべて単一の計算機システム10上で実行する必要は必ずしもない。例えば、ある1つの計算機システムが、CGI（Common Gateway Interface）などの遠隔手続き呼び出しの仕組みを利用することによって、WWWサーバ経由で情報コンテンツ

検索要求を受信した後、ネットワーク上の他のサーバに対して検索処理の実行を委ねるようにしてもよい。

【0121】また、ネットワーク上では、ソフトウェア・プログラムやデータ・コンテンツなどのディストリビューションが行われる（周知）。例えば、本実施例に係るリコmendーション・アプリケーションをネットワーク経由で配信したり、コンテンツ検索のためのキーワード・ベクトルや顧客プロファイルなどのデータ・ファイルをネットワーク経由で配信することが可能である。

【0122】図9には、本実施例に係る情報検索システム10が、情報コンテンツとしての書籍の検索のために使用するデータベースを一覧表示している。但し、これらのデータベースをすべて単一の計算機システム上で蓄積・管理する必要はなく、ネットワーク上の他の1以上のデータベース・サーバやファイル・サーバに分散して配置して、必要に応じて適宜アクセスするようにしてもよい。

【0123】次いで、情報検索システム10上で実行される情報検索及びリコmend情報の生成処理について説明する。以下で説明する処理手順は、例えば、一般的な計算機システムで構成される情報検索システム10上で、本実施例に係るリコmendーション・アプリケーションを実行することで実現される。また、以下の説明では、情報検索システム10は書籍の検索及びリコmendーションに適用されているものとする。

【0124】図10には、本実施例に係るコンテンツ検索の処理手順を概略的に示している。

【0125】情報検索システム10は、コンテンツとしての書籍を検索するために、まずテンポラリ・テーブルを初期化する（ステップS1）。テンポラリ・テーブルには、コンテンツ検索のために使用されるキーワードを登録するための「キーワード・テーブル」、各コンテンツすなわち書籍毎にキーワードの出現頻度（又はキーワード・ベクトル）を登録するための「コンテンツ・キーワード・テーブル」、コンテンツ検索の要求元である顧客のプロファイルを管理する「顧客プロファイル・テーブル」、並びに、本処理結果の戻り値であるリコmend情報を登録するために「推薦コンテンツ・テーブル」が含まれる。

【0126】次いで、コンテンツ・キーワードの作成処理を行う（ステップS2）。コンテンツ・キーワードの作成処理は、図11にさらに詳しく記述しているが、詳細は後述に譲る。

【0127】次いで、顧客プロファイルの作成処理を行う（ステップS3）。コンテンツ・キーワードの作成処理は、図12にさらに詳しく記述しているが、詳細は後述に譲る。

【0128】次いで、推薦コンテンツの選択処理を行う（ステップS4）。本実施例における推薦コンテンツの選択処理は、書籍のリコmend情報の生成に該当する。

前述したように書籍リコメンドには、順マッチング型リコメンデーションと、交差マッチング型リコメンデーション、及び、両者のハイブリッド型リコメンデーションが挙げられる。順マッチング型リコメンデーションの処理手順については図13に、交差マッチング型リコメンデーションの処理手順については図14及び図15に、それぞれ詳しく記述しているが、詳細は後述に譲る。

【0129】図11には、コンテンツ・キーワード作成の処理手順の詳細を、フローチャートの形式で図解している。以下、このフローチャートに従って説明する。

【0130】まず、コンテンツ・テーブルからレコードを1つずつ読み出す（ステップS11）。そして、書籍の説明文をキーワードに分解して品詞を付与する（ステップS12）。

【0131】次いで、品詞が不要品詞リストに含まれているか否かをチェックする（ステップS13）。次いで、キーワードが不要品詞リストに含まれているか否かをチェックする（ステップS14）。

【0132】不要品詞リストには、テキスト・ドキュメントなどのコンテンツ中で頻繁に出現してもあまり意味を持たない単語（例えば、冠詞や、「情報」や「システム」のようにそもそも頻繁に出現する性質の用語など）が含まれている。不要品詞リストに含まれている品詞やキーワードは、キーワードから除外する（ステップS15）。

【0133】次いで、コンテンツ・キーワード・テーブル（書籍キーワード・テーブル）に一致するレコードが既に存在するか否かをチェックする（ステップS16）。一致するレコードが発見された場合には、該レコードの出現頻度を1だけ増分する（ステップS17）。他方、一致するレコードが見つからなかった場合には、コンテンツ・キーワード・テーブルにこのレコードを追加する（ステップS18）。この新規レコードには、出現頻度として1を、重みとしてゼロを与える。

【0134】ステップS13～S18の処理を、すべてのキーワードに対して繰り返し実行する（ステップS19）。また、ステップS11～S19の処理を、すべてのコンテンツ（すなわち書籍）に対して繰り返し実行する（ステップS20）。

【0135】次いで、コンテンツ・キーワード・テーブルからキーワード毎に、レコード数をカウントして読み出す（ステップS21）。

【0136】次いで、読み出したすべてのレコードをキーワード・テーブルに追加する（ステップS22）。このとき、コンテンツ頻度にレコード件数を代入する。

【0137】次いで、コンテンツ・キーワード・テーブルからレコードを1つ読み出す（ステップS23）。

【0138】次いで、キーワード・テーブルから同一キーワードのレコードを読み出す（ステップS24）。

【0139】次いで、重み（＝出現頻度÷log（全コン

テンツ数÷コンテンツ頻度））を計算する（ステップS25）。

【0140】次いで、コンテンツ・キーワード・テーブル中の該当するレコードを、上記の計算した値で更新する（ステップS26）。

【0141】そして、ステップS23～S26の処理をすべてのコンテンツ・キーワードに対して繰り返し実行することにより、コンテンツ・キーワードすなわち書籍キーワードが完成する。

【0142】また、図12には、顧客プロフィール作成の処理手順の詳細を、フローチャートの形式で図解している。以下、このフローチャートに従って説明する。

【0143】まず、リコメンド要求元である顧客の購買履歴テーブルからレコードを1つずつ読み出す（ステップS31）。

【0144】そして、コンテンツ・キーワード・テーブルの中から、コンテンツIDが一致するレコードを1つだけ読み出す（ステップS32）。

【0145】次いで、顧客プロフィール・テーブル中に、顧客ID、カテゴリID、及び、キーワードが一致するレコードが存在するか否かをチェックする（ステップS33）。

【0146】一致するレコードが存在する場合には、この一致レコードの重みに該処理中のレコードの重みを加算する（ステップS34）。

【0147】次いで、顧客プロフィール・テーブルにこのレコードを追加する（ステップS35）。

【0148】上記のステップS32～S35の処理を、コンテンツIDが一致するすべてのレコードに対して実行する（ステップS36）。

【0149】そして、ステップS31～S36の処理を、購買履歴テーブル中のすべてのレコードに対して実行する（ステップS37）。

【0150】また、図13には、順マッチング型リコメンデーションの処理手順の詳細を、フローチャートの形式で図解している。以下、このフローチャートに従って説明する。

【0151】まず、顧客プロフィールから顧客、カテゴリID毎にキーワード・ベクトルを読み出す（ステップS41）。

【0152】次いで、コンテンツ・テーブルからカテゴリIDが一致するコンテンツを1つ読み出す（ステップS42）。

【0153】次いで、コンテンツ・キーワード・テーブルからコンテンツIDが一致するキーワードを読み出す（ステップS43）。

【0154】そして、顧客プロフィール側のキーワード・ベクトルと、コンテンツ側のキーワード・ベクトルを用いて、推薦度を計算する（ステップS44）。顧客プロフィールは、図12に示す処理手順に従って既に作成

されている。

【0155】また、顧客 U_k のカテゴリ C_i におけるキーワード t_j の重みを w'_{jkl} とおき、コンテンツ（書籍） d_i におけるキーワード t_j の重み w_{ij} とおいたとき、顧客 U_k のカテゴリ C_i に対するコンテンツ d_i の推薦度 r_{ikl} は、上記の式〔数5〕によって表される。但し、顧客 U_k のカテゴリ C_i にキーワード t_j が存在しない場合は w'_{jkl} をゼロとして扱い、また、コンテンツ d_i にキーワード t_j が存在しない場合は w_{ij} をゼロとして扱う。

【0156】次いで、推薦コンテンツ・テーブルに新規レコードを追加する（ステップS45）。

【0157】上記のステップS42～S45の処理を、カテゴリIDが一致するすべてのコンテンツすなわち書籍に対して繰り返し実行する（ステップS46）。

【0158】また、上記のステップS41～S46の処理を、顧客プロフィール中のすべてのキーワード・ベクトルに対して繰り返し実行する（ステップS47）。

【0159】次いで、推薦コンテンツ・テーブルから、顧客毎に推薦度が高い順にレコードを読み出す（ステップS48）。

【0160】そして、購買履歴テーブル中に顧客ID及びコンテンツIDの双方が一致するレコードが存在するか否かをチェックする（ステップS49）。一致するレコードは、推薦コンテンツ・テーブルから削除する（ステップS54）。

【0161】さらに、推薦履歴テーブル中に顧客ID及びコンテンツIDの双方が一致するレコードが存在するか否かをチェックする（ステップS50）。一致するレコードは、推薦コンテンツ・テーブルから削除する（ステップS54）。

【0162】次いで、推薦コンテンツを顧客毎にカウントする（ステップS51）。そして、カウントした件数が推薦したいコンテンツ数を越えたか否かをチェックする（ステップS52）。推薦したいコンテンツ数以内であれば、推薦履歴テーブルに当該レコードを追加登録する（ステップS53）。逆に、推薦したいコンテンツ数を越えてしまった場合には、当該レコードを推薦コンテンツ・テーブルから削除する（ステップS54）。

【0163】上記のステップS48～S54の処理を、推薦コンテンツ・テーブル中のすべてのレコードに対して繰り返し実行する（ステップS55）。

【0164】また、図14及び図15には、交差マッチング型リコメンデーションの処理手順の詳細を、フローチャートの形式で図解している。以下、これらのフローチャートに従って説明する。

【0165】まず、リストAを初期化する（ステップS61）。

【0166】そして、顧客テーブルからレコードを1つ読み出す（ステップS62）とともに、購買履歴テーブ

ルから顧客IDするレコードを読み出す（ステップS63）。

【0167】次いで、購買履歴のあるカテゴリIDのリストを作成して、リストAに書き込む（ステップS64）。

【0168】次いで、顧客プロフィールの中から、顧客IDが一致するレコードをカテゴリID毎に読み出す（ステップS65）。

【0169】次いで、コンテンツ・テーブルから以下の条件1を満たすコンテンツを1つ読み出す（ステップS66）。

【0170】条件1：カテゴリIDが顧客プロフィールのカテゴリIDと異なり（カテゴリが交差する）、且つ、カテゴリIDがリストAに含まれること。

【0171】次いで、コンテンツ・キーワード・テーブルの中から、コンテンツIDが一致するキーワード・ベクトルを読み出す（ステップS67）。顧客プロフィールは、図12に示す処理手順に従って既に作成されている。

【0172】次いで、顧客プロフィール側のキーワード・ベクトルと、コンテンツ側のキーワード・ベクトルを用いて、推薦度を計算する（ステップS68）。顧客プロフィールは、図12に示す処理手順に従って既に作成されている。

【0173】また、顧客 U_k のカテゴリ C_i におけるキーワード t_j の重みを w'_{jkl} とおき、コンテンツ（書籍） d_i におけるキーワード t_j の重み w_{ij} とおいたとき、顧客 U_k のカテゴリ C_i に対するコンテンツ d_i の推薦度 r_{ikl} は、上記の式〔数5〕によって表される。但し、顧客 U_k のカテゴリ C_i にキーワード t_j が存在しない場合は w'_{jkl} をゼロとして扱い、また、コンテンツ d_i にキーワード t_j が存在しない場合は w_{ij} をゼロとして扱う。

【0174】次いで、推薦コンテンツ・テーブルの中に、顧客IDとコンテンツIDの双方が一致するレコードが存在するか否かをチェックする（ステップS69）。一致するレコードが存在しない場合には、推薦コンテンツ・テーブルに当該レコードを追加登録する（ステップS72）。

【0175】また、推薦コンテンツ・テーブルの中に、顧客IDとコンテンツIDの双方が一致するレコードが存在する場合には、さらに、当該レコードの推薦度がステップS68で計算した推薦度よりも高いか否かをチェックする（ステップS70）。計算した推薦度よりも高い場合には、当該レコード上の推薦度を、計算した推薦度の値で更新する（ステップS71）。また、計算した推薦度よりも高くない場合には、ステップS71をスキップする。

【0176】上記のステップS66～S72の処理を、カテゴリIDが条件1に合致するすべてのコンテンツに

対して繰り返し実行する（ステップS73）。

【0177】また、上記のステップS65～S73の処理を、該当する顧客のすべての顧客プロフィール・キーワード・ベクトルに対して繰り返し実行する（ステップS74）。

【0178】また、ステップS61～S74の処理を、顧客テーブル中のすべての顧客に対して繰り返し実行する（ステップS75）。

【0179】次いで、推薦コンテンツ・テーブルの中から、顧客毎に推薦度が高い順で1つずつ読み出す（ステップS76）。

【0180】そして、購買履歴テーブル中に顧客ID及びコンテンツIDの双方が一致するレコードが存在するか否かをチェックする（ステップS77）。一致するレコードは、推薦コンテンツ・テーブルから削除する（ステップS82）。

【0181】さらに、推薦履歴テーブル中に顧客ID及びコンテンツIDの双方が一致するレコードが存在するか否かをチェックする（ステップS78）。一致するレコードは、推薦コンテンツ・テーブルから削除する（ステップS82）。

【0182】次いで、推薦コンテンツを顧客毎にカウントする（ステップS79）。そして、カウントした件数が推薦したいコンテンツ数を越えたか否かをチェックする（ステップS80）。推薦したいコンテンツ数以内であれば、推薦履歴テーブルに当該レコードを追加登録する（ステップS81）。逆に、推薦したいコンテンツ数を越えてしまった場合には、当該レコードを推薦コンテンツ・テーブルから削除する（ステップS82）。

【0183】上記のステップS76～S82の処理を、推薦コンテンツ・テーブル中のすべてのレコードに対して繰り返し実行する（ステップS83）。

【0184】《注釈》

[1]: G. Salton and C. Buckley, Term-weighting approaches in automatic text retrieval., Information Processing and Management, 1988, Vol.14, No.5, P.513-523

[2]: Marko Balabanovic, and Yoav Shoham, Content-Based, Collaborative Recommendation., Communications of the ACM, 1997, Vol. 40, No.3, p66-72

[3]: Paul Resnick, Neophytos Iacovou, Mitesh Suchak, et al., GroupLens: An Open Architecture for Collaborative Filtering of Netnews. Proceedings of the Conference on Computer Supported Cooperative Work, 1994, p175-186

[4]: NetPerceptions Inc., Recommendation Engine White Paper. <http://www.netperceptions.com/literature/content/recommendation.pdf>, 2000

[5]: Eric Brill, Rule Based Tagger, <http://www.cs.jhu.edu/~brill/home.html>, 2000

【0185】[追補] 以上、特定の実施例を参照しながら、本発明について詳解してきた。しかしながら、本発明の要旨を逸脱しない範囲で当業者が該実施例の修正や代用を成し得ることは自明である。

【0186】本実施例では、書籍販売をベースに説明したが、他の商品の検索、あるいは商品以外の物品やデータ・コンテンツを探し当てる場合であっても、本発明が同様に実現可能であることは言うまでもない。また、必ずしも、インターネットなどの広域ネットワーク上でのコンテンツ検索に限定されない。例えば、スタンドアロンの計算機システム上で、対話画面上でコンテンツ検索サービスを検索する場合であっても、同様に本発明を適用することかできる。

【0187】要するに、例示という形態で本発明を開示してきたのであり、限定的に解釈されるべきではない。本発明の要旨を判断するためには、冒頭に記載した特許請求の範囲の欄を参酌すべきである。

【0188】

【発明の効果】以上詳記したように、本発明によれば、膨大な情報コンテンツの中からユーザの要求に合致したコンテンツを探し出してリコメンドすることができる、優れた情報検索システム又は方法を提供することができる。

【0189】また、本発明によれば、ユーザの趣味・嗜好や過去の探索履歴・購買履歴などのユーザプロフィール情報に基づいて所望のコンテンツの探索しリコメンドを行うことができる、優れた情報検索システム又は方法を提供することができる。

【0190】また、本発明によれば、インターネット上で商品販売などのビジネスを行う場合において、顧客が期待する商品を探し当ててリコメンドすることができる、優れた情報検索システム又は方法を提供することができる。

【0191】また、本発明によれば、書籍販売などのように、顧客が有益性の他に意外性を求めるような商品を探し当てて顧客にリコメンドすることができる、優れた情報検索システム又は方法を提供することができる。

【0192】本発明に係る情報検索システム及び方法は、コンテンツ分析方式を基調とするが、リコメンド情報の幅を広げる工夫をすることにより、コンテンツ分析方式の欠点を解消することができる。また、書籍販売のように対象コンテンツ数が非常に大きな業種において本発明を適用することで、顧客規模が小さい場合にも、質の高いリコメンド・サービスを提供可能にすることができる。

【図面の簡単な説明】

【図1】本発明に係る情報検索システムがリコメンド情報を生成する手順を示した図である。

【図2】本実施例において仮定される、顧客が実際の書店で興味にあった書籍を探し出すときの書籍購買行動モ

デルを示した図である。

【図3】BISAC Subject Categoryを用いて、顧客の書籍購買カテゴリ数を調べたグラフである。

【図4】顧客プロフィールのキーワード・ベクトル作成処理手順を示した図である。

【図5】書籍毎のキーワード・ベクトルの作成手順を示した図である。

【図6】順マッチング型リコメンデーションの処理手順を示した図である。

【図7】交差マッチング型リコメンデーションの処理手順を示した図である。

【図8】本発明の実施に供される情報検索システム10のハードウェア構成を模式的に示した図である。

【図9】本実施例に係る情報検索システム10が、情報コンテンツとしての書籍の検索のために使用するデータベースを一覧表示した図である。

【図10】本実施例に係るコンテンツ検索の処理手順を概略的に示したフローチャートである。

【図11】コンテンツ・キーワード作成の処理手順を詳細に示したフローチャートである。

【図12】顧客プロフィール作成の処理手順を詳細に示したフローチャートである。

【図13】順マッチング型リコメンデーションの処理手順を詳細に示したフローチャートである。

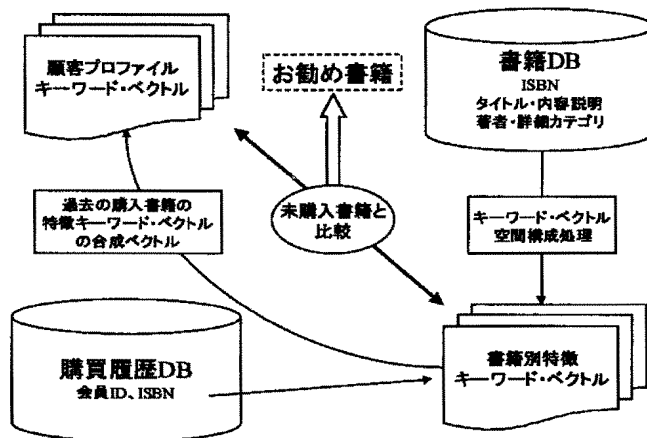
【図14】交差マッチング型リコメンデーションの処理手順（前半）を詳細に示したフローチャートである。

【図15】交差マッチング型リコメンデーションの処理手順（後半）を詳細に示したフローチャートである。

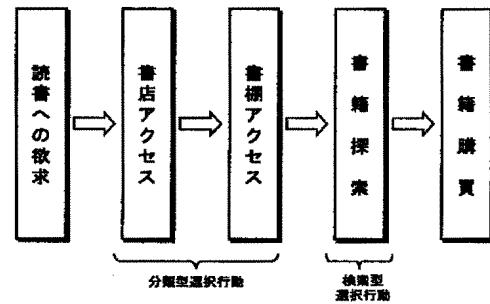
【符号の説明】

- 10…情報検索システム
- 11…CPU
- 12…RAM
- 13…ROM
- 14…ディスプレイ・インターフェース
- 15…入力機器インターフェース
- 16…ネットワーク・インターフェース
- 17…外部機器インターフェース
- 19…バス
- 21…ディスプレイ
- 22…キーボード
- 23…マウス
- 24…ハード・ディスク装置
- 25…メディア・ドライブ

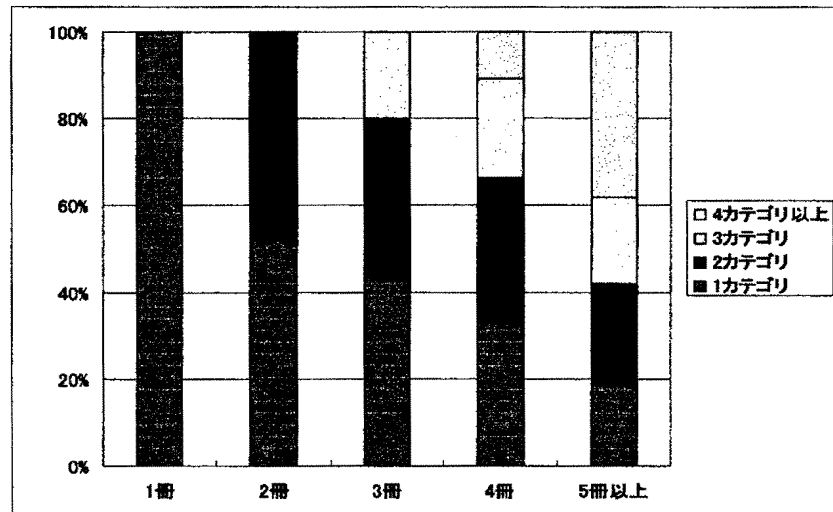
【図1】



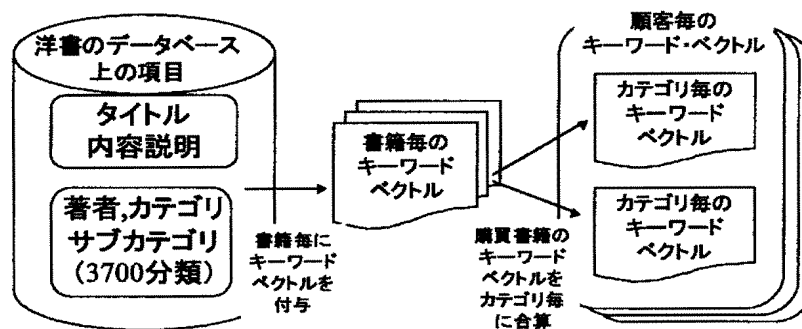
【図2】



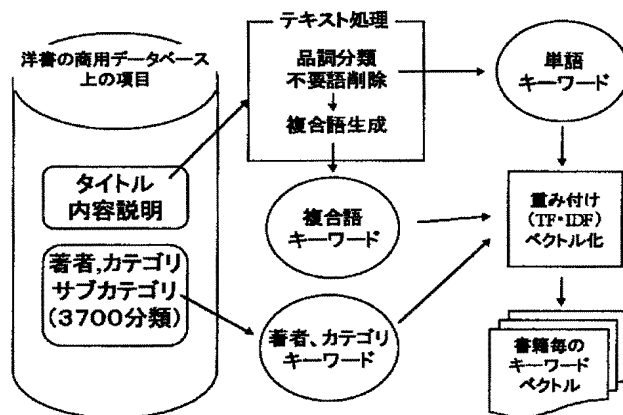
【図3】



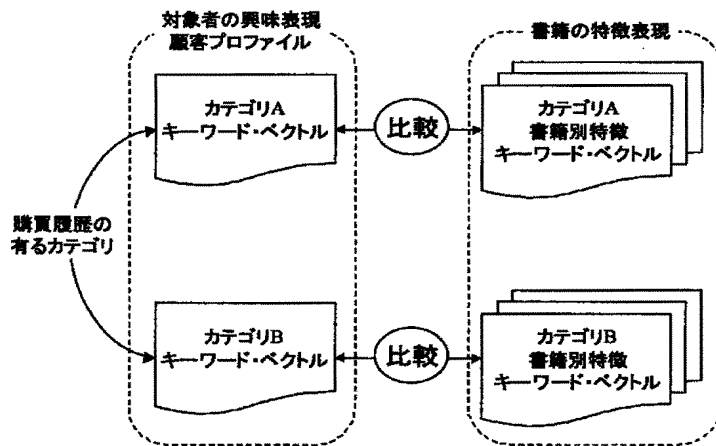
【図4】



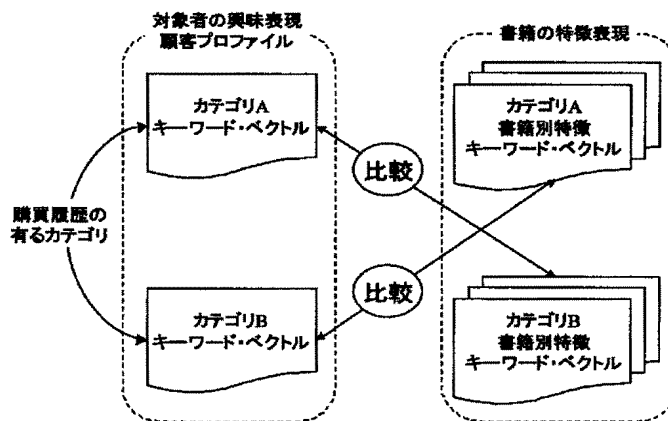
【図5】



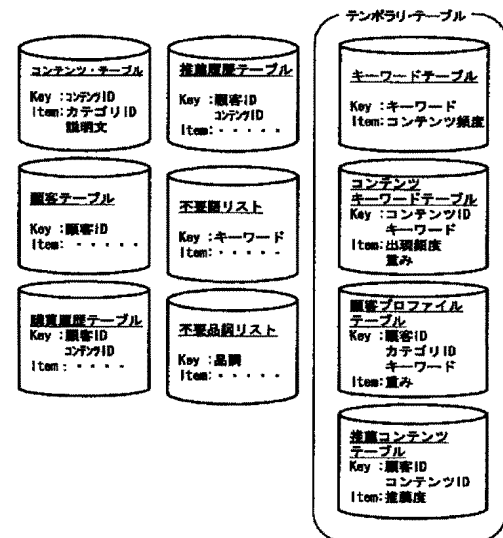
【図6】



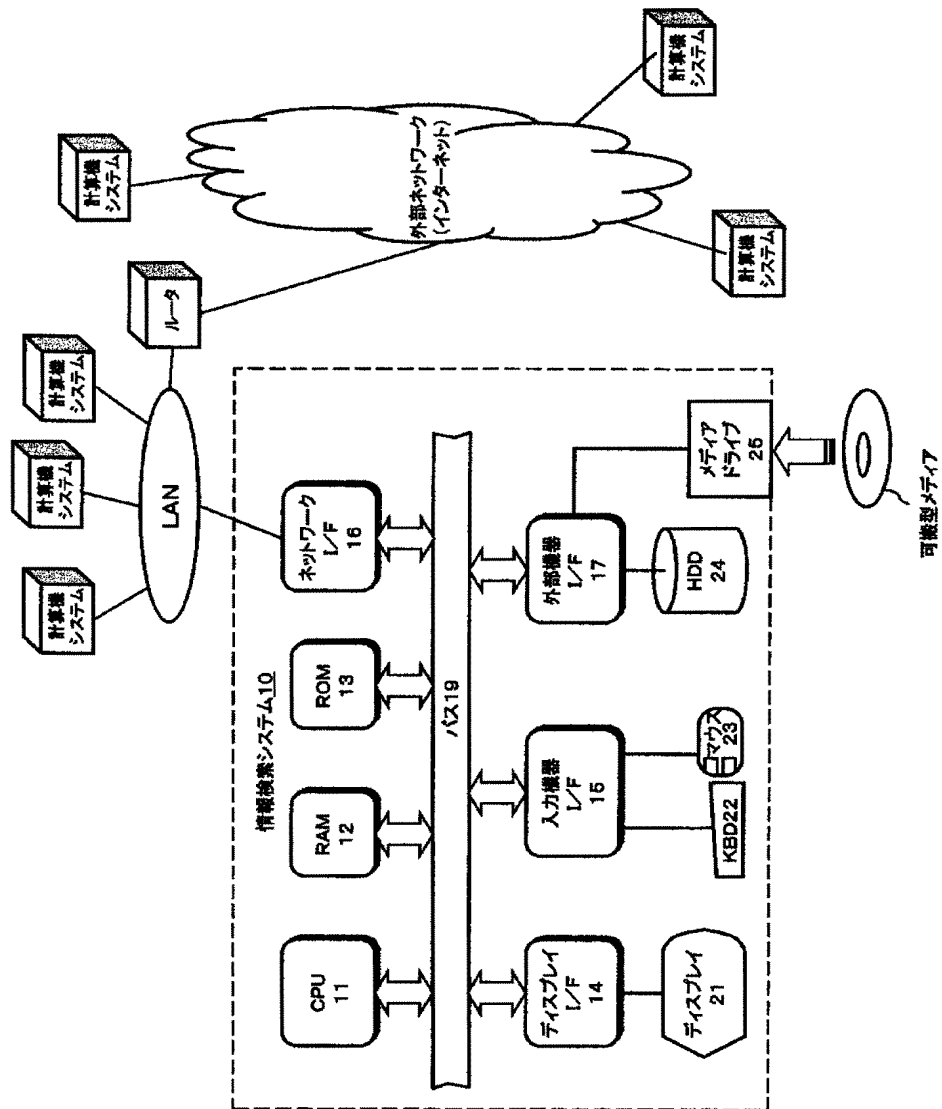
【図7】



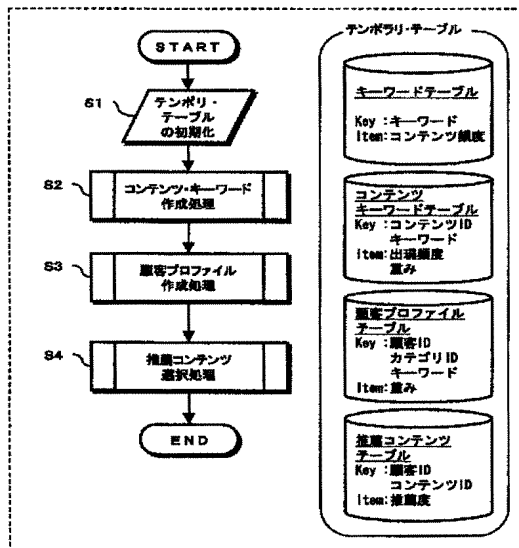
【図9】



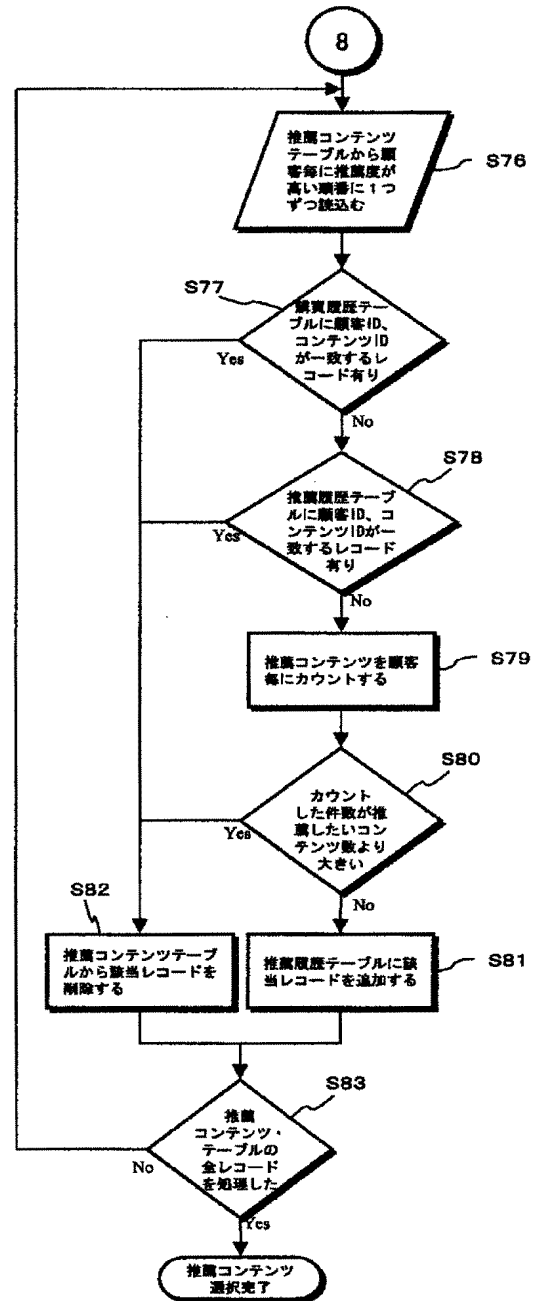
【図8】



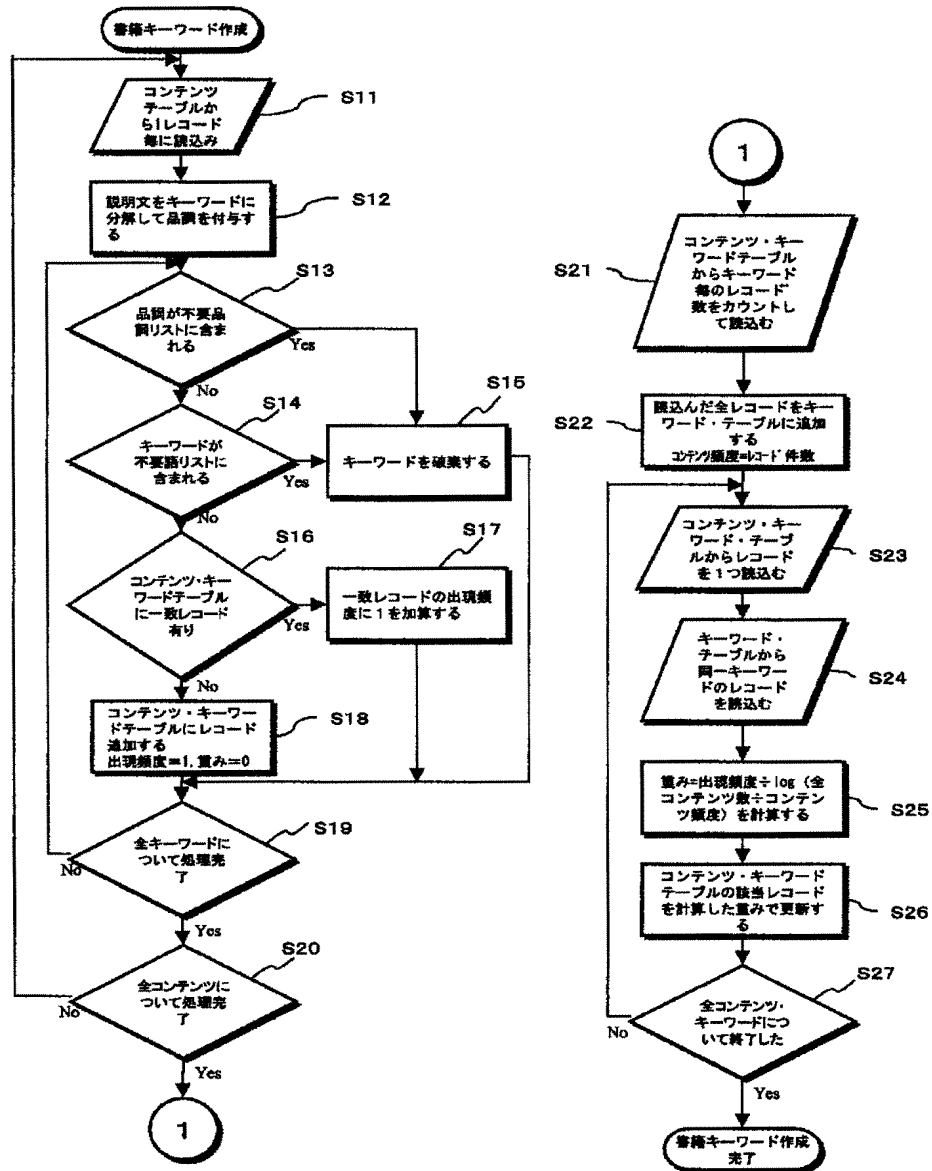
【図10】



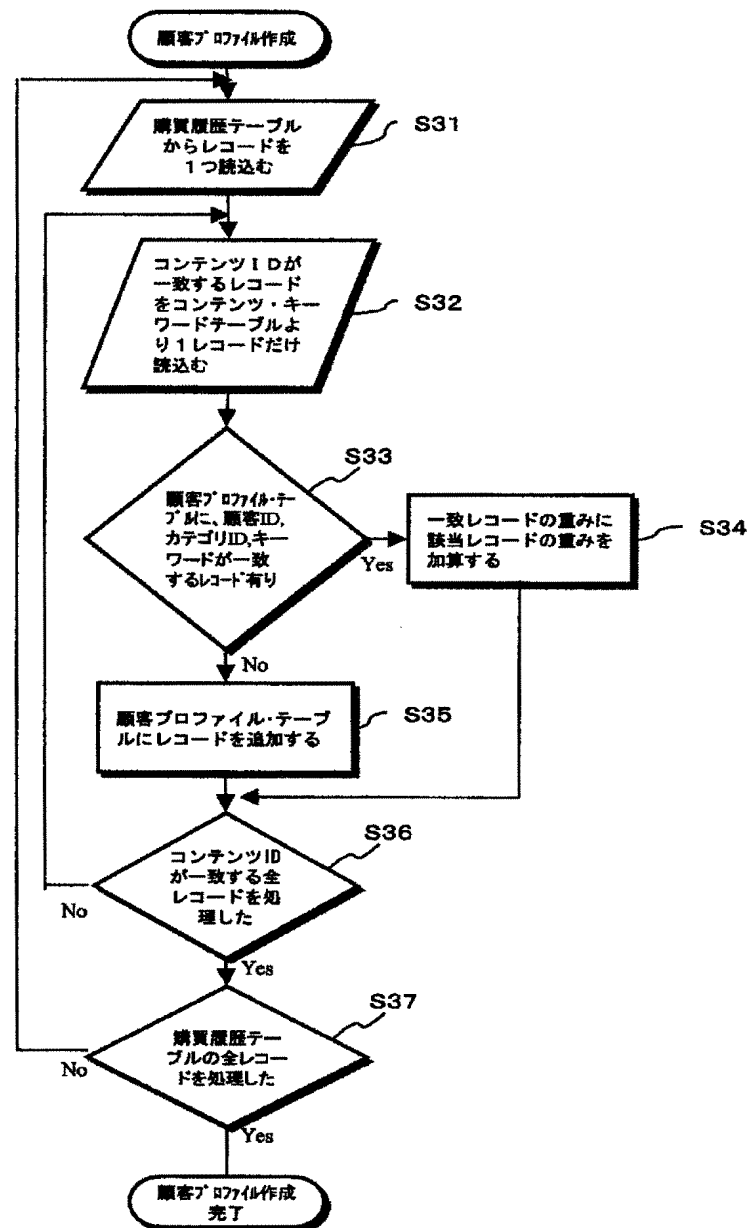
【図15】



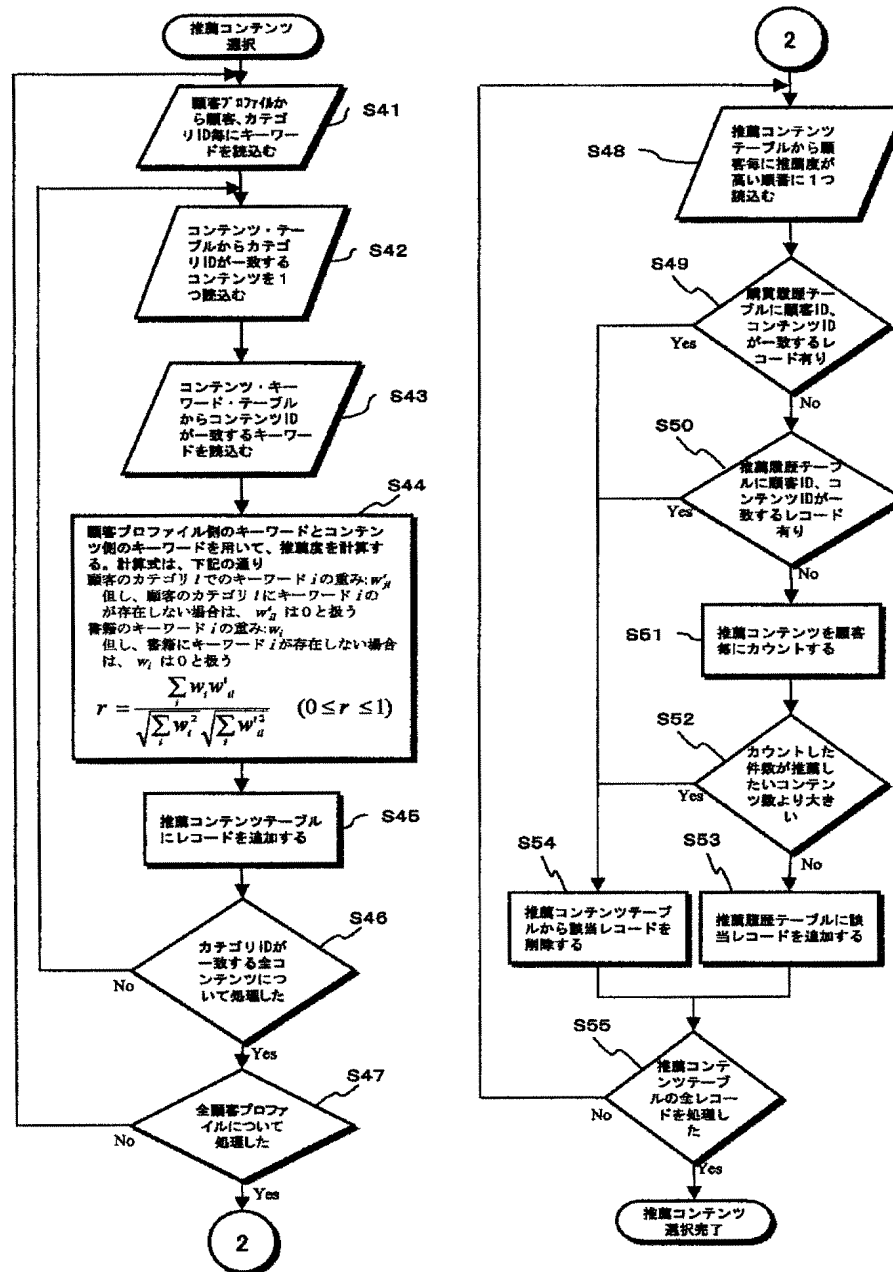
【図11】



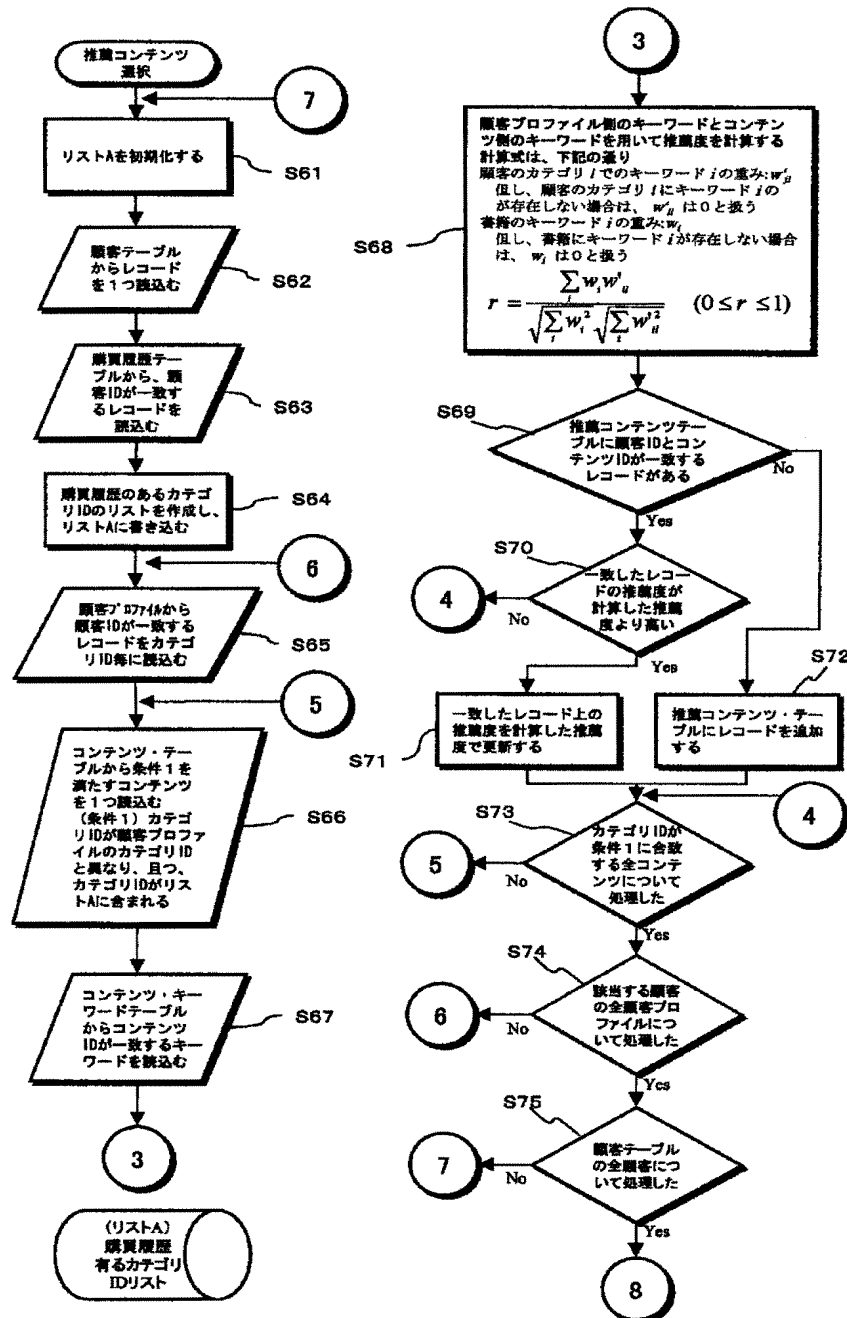
【図12】



【図13】



【図14】



フロントページの続き

Fターム(参考) 5B075 NK46 NR12 PP02 PP03 PQ02
PR04 PR06 QM08 UU11